STATISTICAL COMPARISONS FOR NONLINEAR CURVES AND SURFACES

Shi Zhao

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Biostatistics,

Indiana University

August 2018

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Wanzhu Tu, Ph.D., Chair

Doctoral Committee

Giorgos Bakoyannis, Ph.D.

Spencer Lourens, Ph.D.

May 31, 2018

Yiqing Song, M.D., Sc.D.

DEDICATION

To My Loving Family.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor **Wanzhu Tu**, for his valuable guidance, enthusiastic inspiration, and continuous support of my PhD research. Without him, my research journey wouldn't have been so memorable and pleasant. I am deeply grateful to Dr. **George Bakoyannis** for his help on the theoretical proofing as well as his insightful ideas in many aspects of my dissertation structure. I am thankful to Dr. **Spencer Lourens** for his sophisticated computing skills and ever-friendly nature. His advices facilitated my R package and Shiny interface development. It's also my privilege to offer my gratitude to Professor **Yiqing Song**, who is also my minor advisor, for his intellectual vigor and supports.

In addition to my committee members, I would like to give my special thanks to Professor **Barry Katz**, who supported me as a research assistantship in a nationwide grant for two years of my PhD study. Finally, I wish to express my gratitude to all Biostatistics program members, including faculty, staff, and other student colleagues for offering such a friendly academic environment and helping me in all aspects during my work here.

Shi Zhao

STATISTICAL COMPARISONS FOR NONLINEAR CURVES AND SURFACES

Estimation of nonlinear curves and surfaces has long been the focus of semiparametric and nonparametric regression. The advances in related model fitting methodology have greatly enhanced the analyst's modeling flexibility and have led to scientific discoveries that would be otherwise missed by the traditional linear model analysis. What has been less forthcoming are the testing methods concerning nonlinear functions, particularly for comparisons of curves and surfaces. Few of the existing methods are carefully disseminated, and most of these methods are subject to important limitations. In the implementation, few off-the-shelf computational tools have been developed with syntax similar to the commonly used model fitting packages, and thus are less accessible to practical data analysts. In this dissertation, I reviewed and tested the existing methods for nonlinear function comparison, examined their operational characteristics. Some theoretical justifications were provided for the new testing procedures. Real data examples were included illustrating the use of the newly developed software. A new R package and a more user-friendly interface were created for enhanced accessibility.

Wanzhu Tu, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Linear regression has been a workhorse of practical data analysis in much of the 20th century (Draper and Smith, 1998). While parametric linear and generalized linear regression models continue to dominate today's analytical landscape, there is an increasing awareness that in many settings, especially in complex biological studies, few effects are truly linear, or can be adequately described by analyst-supplied parametric functions (Green and Silverman, 1994). In those situations, forcing a parametric model amounts to a form of model misspecification, which result in erroneous estimation and inference.

Efforts to overcome this limitation have given rise to nonparametric and semiparametric regression methods, including local polynomial models (Fan and Gijbels, 1996), wavelet methods (Ogden, 1996), smoothing splines (Gu, 2013; Wahba, 1990), and various penalized spline models (de Boor, 2001; Eilers and Marx, 1996; Eubank, 1999). By expressing the effects of individual explanatory variables as smooth functions, Hastie and Tibshirani's generalized additive models (GAM) further extend the boundary of nonparametric regression (Hastie and Tibshirani, 1990). Bridging the gap between parametric and nonparametric regression models, Ruppert, Wand, and Carroll's semiparametric regression models were based on penalized splines (Ruppert et al., 2003). Through a mixed effect model expression, these semiparametric models have greatly influenced the modeling of nonlinear effects in practical data analysis. Surveying the recent biomed-

ical literature, we see a rapid increase in the use of these models mostly along the lines described by Ruppert et al.'s work.

Much of the methodological development of nonparametric and semiparametric regression in the last two decades has been on the estimation of nonlinear effects. There is a sizable literature on the estimation of nonlinear functions using various nonparametric techniques. Among the available computational packages, Hastie's `gam` and Wood's `mgcv` and `gamm4` are frequently used in practical data analysis (Hastie, 2006; Wood, 2008; Wood, 2017). Gu's (2014) `gss` is also a popular choice. These well designed software packages have enhanced the analyst's toolset for discovering and depicting nonlinear relations. In our own experience, these different methods often generate similar nonlinear function estimates in real data applications. As a result, the choice of smoothing methods is often less consequential, driven mainly by considerations of implementational convenience, software availability, and analysts' personal preference. Estimation, although important, is only the first step in an exploration, however, and statistical inference remains the ultimate analytical objective. It is towards this end that statistical methodology has not been able to keep up with the demand of science (Wood, 2018).

Although nonlinear curves and surface estimation saw its mosst rapid development in the 1990s, major estimation methods were put forward much earlier, including kernel based (Nan et al., 1964), spline based (B-splines (de Boor, 2001; Watson, 1964), Wavelets (Hart, 1997), Fourier-expansion (Hart, 1997)) and penalty based methods (Green and Silverman, 1994). In spite of the increasingly wide application of smoothing regression, testing methods concerning nonlinear functions, particularly comparisons of curves and surfaces across groups, remained less studied.

Until now, only a few publications have studied comparisons of smoothing functions. Among these, Fan et al. (1996, 1998) constructed a test of significance between curves based on the adaptive Neyman test and the wavelet thresholding technique (Fan, 1996; Fan and Lin, 1998). Dette and Neumeyer (2001, 2003) developed three testing procedures on the equality of k regression curves from independent samples in a kernel-based setting (Dette and Neumeyer, 2001). Zhang and Lin (2003) considered testing nonparametric functions in the semiparametric additive mixed models, and they constructed a test statistic following a scaled χ^2 distribution under the null hypothesis (Zhang and Lin, 2003). Bowman (2006) proposed a surface testing method using χ^2-approximation with kernel smoothing (Bowman, 2006). More recently, Wang (2010) extended Dette and Neumeyer's L^2-distance method to surface comparison for both homoscedastic and heteroscedastic models (Wang and Ye, 2010). The testing method proposed by Zhang and Lin was the only one based on a semiparametric modeling technique; the method, however, is only applicable when the values of explanatory variable(s) for the nonparametric function were the same across groups. These limitations have impeded the application of the aforementioned methods.

In this dissertation, I proposed two extensions to the existing methods. The first testing method was built on penalized semiparametric estimation, and it used a wild bootstrap procedure for comparing nonlinear curves and surfaces. The method was developed for analysis of cross-sectional data. The second method was for the analysis of clustered data and is essentially a mixed effect model extension of the first method. Collectively, the two methods provide practical solutions to a broad class of inference problems involving comparisons of nonlinear functions. In their accommodation of co-

variates and data correlations, the methods were less restrictive than the existing methods.

This thesis starts with a comprehensive review of the recent literature on the field of comparison of nonparametric and semiparametric regression functions. In Chapter 2, I describe the first hypothesis testing method, which was based on an L^2-distance of pointwise semiparametric estimated regression functions. I used a wild bootstrap procedure to approximate the critical values of the test statistic. Under the null hypothesis, I conducted extensive simulation studies to examine the performance of the proposed method, including both testing power and Type I error rate. In Chapter 3, I provide theoretical and numerical justifications for the new method. In Chapter 4, I extend the method to correlated data and provide corresponding simulation results. Finally, in Chapter 5, I present an R package `gamm4.test` and an interactive R-Shiny interface for the testing procedures.

1.1 Existing Nonparametric Statistical Methods for Nonlinear Curve and Surface Comparison

1.1.1 Nonlinear curve comparison

Fan's Wavelet transformation testing method

Fan et al (1996, 1998) studied two test statistics based on the adaptive Neyman tests and the wavelet thresholding. They considered testing the hypothesis of two cumulative distribution functions $H_0 : G(x) = G_0(x)$ vs $H_1 : G(x) \neq G_0(x)$, where $X_1, \ldots, X_n$ were n iid sample. Due to the limitation of insufficient power of the Kolmogorov-Smirnov test and the Cramer-Von Mises test when the densities contained subtle local features, Fan

proposed to first conduct a Fourier transformation using G_0 so that the test became $H_0 : G(x) =$ uniform vs $H_1 : G(x) \neq$ uniform, which was also equivalent to test the Fourier coefficients $H_0 : \theta_j = 0, j = 1,2,\ldots$ vs $H_1 :$ at least one of $\theta_j \neq 0$.

The explicit adaptive Neyman tests and wavelet thresholding tests by Fan are illustrated as follows. Let $\mathbf{Z} \sim N(\theta, \mathbf{I}_n)$ be an n-dimentional random vector. To test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$, the adaptive Neyman test is to test only the first m components of the large absolute values of θ, where m is estimated from

$$\hat{m} = argmax_{m:1\leq m\leq n}\{m^{-1/2}\sum_{j=1}^{m}(Z_j^2 - 1)\},$$

for $j = 1,\ldots,m$. This method circumvents the problem of testing in a high dimensional space, where large accumulated stochastic noise and decreased spower plagues the test. The resulting adaptive Neyman test statistic takes the form

$$T_{AN} = \sqrt{2loglogn}(\sqrt{2\hat{m}})^{-1}\sum_{j=1}^{\hat{m}}(Z_j^2 - 1) - \{2loglogn + 0.5loglogloglogn - 0.5log(4\pi)\}.$$

T_{AN} could be compared with the asymptotic extreme value distribution for the purpose of hypothesis testing. Wavelet thresholding test statistic is defined based on a wavelet transform of the observation vector $\mathbf{Z}$. The test statistic takes the form $T_H = \sum_{j=1}^{n} Z_j^2 I(|Z_j| > \delta)$, where $\delta > 0$ is a thresholding parameter. They recommended $\delta = \sqrt{2log(n/log^2/n)}$ for better power and accuracy of T_H approximation. Other δ's can also be used if power improvement is needed or a data-dependent thresholding parameter is preferred. Fan showed that T_H followed an asymptotically normal distribution,

hence a test could be constructed by comparing standardized T_H with standard normal distribution.

To compare two sets of curves $Y_{ij}(x) = g_i(x) + \epsilon_{ij}(x)$, where $i = 1,2$, $j = 1,\ldots,n_i$, $\epsilon_{ij}(x) \sim N(0,\sigma_i^2(x))$, one tests the hypothesis $H_0 : g_1 = g_2$ vs $H_1 : g_1 \neq g_2$. Fan et al. chose to use the standardized difference of summarized curves

$$Z(x) = \{n_1^{-1}\hat{\sigma}_1^2(x) + n_2^{-1}\hat{\sigma}_2^2(x)\}^{-1/2}\{\bar{Y}_1(x) - \bar{Y}_2(x)\}$$

for hypothesis testing, where $\bar{Y}_1(x)$, $\bar{Y}_2(x)$ are respectively the mean of Y_{1i} and Y_{2j} at each x; and $\hat{\sigma}_1(x)$, $\hat{\sigma}_2(x)$ are the estimated standard deviations of the individual sets. They showed that $Z(x)$ followed an asymptotically normal distribution, $N(d(x),1)$, with $d(x) \approx \{n_1^{-1}\sigma_1^2(x) + n_2^{-1}\sigma_2^2(x)\}^{-1/2}\{g_1(x) - g_2(x)\}$. Accordingly, an adaptive Neyman test or Fourier transform thresholding test could be applied to test the vector $Z(x)$ for comparing two sets of curves.

Fan (1996) showed through simulation that when the curves were smooth, the adaptive Neyman test could be used; otherwise, the wavelet thresholding test performed better. The adaptive Neyman test could be extended to compare multiple curves, however the wavelet thresholding test has not been extended to multiple curves testing as a good thresholding parameter for wavelet transform is difficult to define in such situations.

One strength of the adaptive Neyman test and the wavelet thresholding test is its ability to detect local characteristics and global alternations. For instance, these methods are well-suited for detecting sharp peaks in spectral density or densitometric tracings. The Fourier transformation contains high-frequency components or local

characteristics of a data set, in contrast to other popular testing procedures (such as those based on splines), and only uses information contained in the low frequencies, so that the analyst could analyze the signal through the statistical properties of the empirical Fourier coefficients. Despite the sensitivity to local features, applications of the two testing methods are limited as these tests can only be used when the two groups have repeated measurements at the same points of independent variable; otherwise the standardized difference of summarized curves $Z(x)$ cannot be constructed. However, in cross-sectional data analyses, most applications of the two group comparisons have different x values, which render the testing methods inoperable.

Young and Bowman's Nonparametric Analysis of Covariance (ANCOVA)

Young and Bowman (1995) described a method for testing of the equality of two or more smooth curves, under the model $Y_{ij} = g_i(X_{ij}) + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2)$ for $i = 1, 2, \ldots, k$, $j = 1, \ldots, n_i$. They considered the test under a homoscedastic assumption that the error variance would remain constant across all k groups. To test $H_0 : g_1 = g_2 = \cdots = g_k$ vs $H_1 : g_i \neq g_j$ for some $i, j \in \{1, \ldots, k\}$, they used a kernel-based smoothing method to approximate g_i. Assuming h_i is the bandwidth for the ith regression function, they considered

$$\hat{g}_i(x) = \frac{\sum_{j=1}^{n_i} K((x - x_{ij})/h_i) y_{ij}}{\sum_{j=1}^{n_i} K((x - x_{ij})/h_i)} \tag{1.1}$$

as the Nadaraya-Watson estimator of g_i.

Under the null hypothesis, they obtained a common regression function by combining data from all k groups

$$\hat{g}(x) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K((x-x_{ij})/h)y_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K((x-x_{ij})/h)}, \tag{1.2}$$

where $h_i = h$.

Therefore, the resultant test statistic is analogous to the one-way analysis of variance,

$$T_1 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\hat{g}(x_{ij}) - \hat{g}_i(x_{ij})]^2}{\hat{\sigma}^2}, \tag{1.3}$$

where $\hat{g}_i$ and $\hat{g}$ are the kernel-based curve estimator. The variance σ_i^2 can be estimated as

$$\hat{\sigma}_i^2 = \frac{1}{2(n_i-1)}\sum_{j=1}^{n_i-1}(y_{i,[j+1]} - y_{i,[j]})^2.$$

A pooled estimate of σ^2 is $\hat{\sigma}^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i-1)\hat{\sigma}_i^2$, where $N = \sum_{i=1}^{k} n_i$.

For examining the distribution of the test statistic T_1, Young and Bowman argued that since the fitted values for $\hat{g}_i$ could be written as $\hat{\mathbf{g}}_i = \mathbf{S}_i\mathbf{y}_i$, where $\mathbf{S}_i$ was an $n_i \times n_i$ matrix, the entire collection of these individual fitted $\hat{\mathbf{g}}_i$ could be represented as $\hat{\mathbf{g}} = \mathbf{S}_d\mathbf{y}$, with $\mathbf{S}_d$ being an $n \times n$ matrix. The numerator of T_1 is $\mathbf{y}^{\mathbf{T}}[\mathbf{S_d} - \mathbf{S_s}]^{\mathbf{T}}[\mathbf{S_d} - \mathbf{S_s}]\mathbf{y} = \epsilon^{\mathbf{T}}[\mathbf{S_d} - \mathbf{S_s}]^{\mathbf{T}}[\mathbf{S_d} - \mathbf{S_s}]\epsilon$. Additionally, $E(\hat{\sigma}^2)$ could be approximated as $\epsilon^{\mathbf{T}}\mathbf{B}\epsilon$, where $\mathbf{B}$ is a symmetric matrix. Consequently, T_1 is a ratio of quadratic forms, which is analogous to the F-tests in linear models. The calculation of p could be completed by matching the

first three moments of the test statistic with those of a shifted and scaled χ^2 distribution ($a\chi^2_b + c$) under H_0.

While the derivation of this test is easily understood and its implementation straightforward, the equal variance assumption can be overly restrictive. Simulation studies have revealed a number of weaknesses. First, when the underlying relationship is linear, the estimate of σ^2 may not be accurate. Additionally, when the explanatory variables x_{ij} have different values among the groups, the power of the test decreases dramatically because the bias cannot be canceled out under H_0. The test statistic has been extended to situations of surface comparison (Young and Bowman, 1995)).

Dette and Neumeyer's three tests using kernel-based estimators

Dette and Neumeyer (2001) proposed three kernel-based testing methods. Writing the curves as $Y_{ij} = g_i(x_{ij}) + \epsilon_{ij}(x_{ij})$, where $i = 1,2,\ldots,k$, $j = 1,\ldots,n_i$, $\epsilon_{ij}(x_{ij}) \sim N(0,\sigma_i^2(x))$, they aimed at testing $H_0 : g_1 = g_2 = \cdots = g_k$ vs $H_1 : g_i \neq g_j$ for some $i, j \in \{1,\ldots,k\}$, under the following conditions: (1) The variances $\sigma_i^2(.)$ are continuous functions; (2) the design points x_{ij} satisfy $\int_0^{x_{ij}} r_i(x)dx = \frac{j}{n_j}$, where $j = 1,\ldots,n_i, i = 1,\ldots,k$, and r_i is a density function; (3) the regression functions are sufficiently smooth, i.e., ≥ 2 times continuously differentiable in the supporting space. The Nadaraya-Watson estimators $\hat{g}_i$ and $\hat{g}$ remain the same as defined in equation (1.1) and (1.2).

The first test statistic T_2 compares the group-specific error variances against that of the combined sample, in a way that is analogous to the one-way ANOVA. Let

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \hat{g}_i(x_{ij}))^2$$

denote the estimated variance of the ith sample and

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \hat{g}(x_{ij}))^2$$

be the variance for the pooled sample.

$$T_2 = \hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{k} n_i \hat{\sigma}_i^2. \tag{1.4}$$

The second test statistic directly assesses the distance between the group-specific curves and the common curve for all groups, at the observed design points x_{ij}, as introduced by Young and Bowman in equation (1.3),

$$T_3 = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\hat{g}(x_{ij}) - \hat{g}_i(x_{ij})]^2. \tag{1.5}$$

In contrast to the comparison of the residual sum of squares in T_2, the new test statistic T_3 compares the curves through the fitted values.

The third test statistic is a summarized distance based on all pairwise comparisons of the estimated individual curves.

$$T_4 = \sum_{i=1}^{k}\sum_{j=1}^{i-1}\int [\hat{g}_i(x) - \hat{g}_j(x)]^2 w_{ij}(x)dx, \tag{1.6}$$

where $w_{ij}(.)$ are positive weight functions.

The asymptotic normality of all three statistics under H_0 and fixed alternatives with different rates has been demonstrated. In addition, they have shown that the asymptotic variance of T_2 is greater or equal to the other two test statistics. However, as the

speed of convergence to normal distribution under the null hypothesis is typically slow for small to moderate sample sizes, the bias always has to be taken into account. No universal superiority for one of these methods can be established. The investigators therefore recommended a wild bootstrap version of the test when studying finite samples (Wu,1986).

The above tests were later extended for comparison of two regression curves with different design points and heteroscedasticity. The new procedure was applicable in the case of different design points and heteroscedasticity. Under similar regularity assumptions, they showed that the two marked empirical processes converged to a centered Gaussian process at a rate of $N^{-1/2}$ under the null, while under the alternative, the mean of the two processes did not go to zero (Dette and Neumeyer, 2001). A test was then constructed based on functions of these empirical processes, such as integration of the squared residual process or supremum of the absolute residuals. For finite samples, they proposed to use test statistics of the supremum of absolute marked empirical process and to apply a wild bootstrap procedure. However, in the simulation studies, these tests did not show enough sensitivity when the regression curves were close and when the sample sizes were moderate.

Zhang and Lin's χ^2 approximation in the setting of semiparametric additive model

Zhang and Lin (2000) considered testing the equivalence of two nonparametric functions. Later, Zhang and Lin (2003) described a test within the framework of additive mixed models,

$$Y_{ijl} = g_i(x_{ijl}) + \mathbf{s}_{ijl}^T \alpha_i + \mathbf{Z}_{ijl}^T \mathbf{b}_{ij}, \tag{1.7}$$

where Y_{ijl} represents the response variable for the ith group ($i = 1,2$), jth cluster ($j = 1,\ldots,n_i$) and kth observation ($k = 1,\ldots,q_{ij}$), x_{ijl} denotes an explanatory variable, $g_1(.)$ and $g_2(.)$ represent the nonparametric functions of two groups, s_{ijl} is a vector of other associated fixed effects, and Z_{ijl} is a vector of random effects.

Let $[T_1, T_2]$ be an interval that specifies the range of continuous predictor x for both groups. To test the hypothesis $H_0 : g_1 = g_2$ vs. $H_1 : g_1 \neq g_2$, the authors suggested the following test statistic

$$G\{\hat{g}_1(x), \hat{g}_2(x)\} = \int_{T_1}^{T_2} \{[\hat{g}_1(x) - \hat{g}_2(x)]^2\} dx, \tag{1.8}$$

where $\hat{g}_1$ and $\hat{g}_2$ were obtained by maximizing the penalized log-likelihood function. The penalized likelihood under the semiparametric additive mixed model for an individual group was $l(g_i, \alpha_i; \mathbf{y}) - \frac{\lambda_i}{2} \int [g_i''(x)]^2 dx$, where λ_i was the smoothing parameter controlling the goodness of fit of the model and roughness of function $g_i(x)$.

As G in Equation (1.8) could be written as a quadratic function of $\mathbf{y}$, Zhang et al. approximated the distribution of $G\{\hat{g}_1(x), \hat{g}_2(x)\}$ with a scaled χ^2 distribution using the moment matching technique (Zhang et al., 2000). To illustrate, they assumed the random effects b_{ij} to be independent and to follow a normal distribution $N(0, D_0(\theta))$, where θ is a vector of variance components. Let λ_i and θ_i be the smoothing parameter and the variance components under individual models of group i. There exists a vector function $\mathbf{c}_i$ such that $\hat{g}_i$ can be written as $\hat{g}_i(x) = \mathbf{c}_i^T(x)\mathbf{y}_i$. Let $\mathbf{c}(x) = [\mathbf{c}_1(x)^T, -\mathbf{c}_2(x)^T]^T$ and $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. It follows that the test statistic G can be written as a quadratic function of $\mathbf{y}$, $G(\mathbf{y}_1, \mathbf{y}_2) = \int_{T_2}^{T_1} \mathbf{y}^T \mathbf{c}(x)\mathbf{c}^T(x)\mathbf{y} dx = \mathbf{y}^T \mathbf{C} \mathbf{y}$, where $\mathbf{C} = \int_{T_2}^{T_1} \mathbf{c}(x)\mathbf{c}^T(x) dx$. Zhang et al. approximated G distribution by a scaled chi-square $\kappa\chi_v^2$, where the scale parameter κ

and the degrees of freedom υ were then calculated by matching the approximate mean and variance of $G\{\hat{g}_1(.), \hat{g}_2(.)\}$ under H_0. Let $\mathbf{E_0}$ and $\mathbf{V_0}$ be the mean and variance of $\mathbf{y}$ under H_0, then the approximate mean e and variance ψ of $G\{\hat{g}_1(.), \hat{g}_2(.)\}$ under H_0 can be calculated as $e = \mathbf{E}_0^T\mathbf{C}\mathbf{E}_0 + tr(\mathbf{C}\mathbf{V}_0), \psi = 2tr(\mathbf{C}\mathbf{V}_0)^2 + 4\mathbf{E}_0^T\mathbf{C}\mathbf{V}_0\mathbf{C}\mathbf{E}_0$, where the unknown parameters are replaced by their maximum penalized likelihood estimators obtained under H_0. Equating e and ψ to the mean and variance of $\kappa\chi^2_\upsilon$ provided $\kappa = \psi/(2e)$ and $\upsilon = 2e^2/\psi$. By defining $\chi^2_{obs} = G_{obs}/\kappa$, where G_{obs} denotes the observed value of G, the approximate p-value for the test statistic $G\{\hat{g}_1(.), \hat{g}_2(.)\}$ is given by $P(\chi^2_{2e/\psi} > \chi^2_{obs})$. To improve the approximation of the distribution of the test statistic $G\{\hat{g}_1(.), \hat{g}_2(.)\}$, one can use higher moments in matching with a shifted and scaled χ^2 distribution, similar to the p value calculation proposed by Young and Bowman. Zhang et al.(2003) also extended this result to non-Gaussian data using a generalized semiparametric additive mixed model.

The test statistic proposed by Zhang et al. is similar to the one from Young and Bowman (1995) in Section 1.1.1; however, Young and Bowman estimated the nonparametric functions using the kernel method. Also, the test statistic (1.8) is equivalent to (1.6) by choosing $w_{ij}(\cdot)$ equal to $f(x)$. When the two groups have the same values of x_{jk} and $\mathbf{s}_{jk}$, the bias in the smoothing spline estimates $\hat{g}_1$ and $\hat{g}_2$ is canceled out under H_0. In situations where the two groups have different values of $(x_{jk}, \mathbf{s}_{jk})$ and (θ, λ), the biases in $\hat{g}_1$ and $\hat{g}_2$ are only partially canceled under H_0. The consequential testing biases were shown in our simulation studies in Section 4.1 Table 4.2. One other major limitation of this method is that the two groups are required to have the same sample size in order to implement the scaled chi-square test algorithm.

1.1.2 Surface comparison

Current surface comparison methods were all generalized from nonlinear curve comparisons. Bowman (2006) adapted the χ^2 approximation of the ANOVA-type test statistic which had been investigated in the univariate case in Young and Bowman (1995). Wang et al. (2011), on the other hand, extended the work of Dette and Neumeyer's (2001, 2003) kernel based nonparametric curve comparison to a surface comparison in company with a wild bootstrap procedure.

Bowman's nonparametric surface comparison method

Suppose we perform a k group comparison, with $(x_{1i},\ x_{2i})$ as independent variables. A model can be formulated as $Y_{ij} = g_i(x_{1ij}, x_{2ij}) + \epsilon_{ij}$, where $i = 1,\ldots,k, \quad j = 1,\ldots,n_i$. We are interested in testing the equality of the mean functions; that is, $H_0 : g_1 = g_2 = \cdots = g_k$ vs. $H_1 : g_i \neq g_j$ for some $i, j \in (1,\ldots,k)$. For the kernel-based method, the conditional expectation of Y relative to $\mathbf{X}$ could be written in $E(Y|\mathbf{X} = \mathbf{x}) = g(\mathbf{x})$. If we denote $\mathbf{H}$ as a bandwidth matrix which is symmetric positive-definite and $det(\mathbf{H})$ as the determinant of the matrix $\mathbf{H}$, the multivariate Nadaraya-Watson estimator of the ith regression function $g_i(\mathbf{x})$ becomes

$$\tilde{g}_{\mathbf{H}i}(\mathbf{x}) = \frac{\sum_{j=1}^{n_i} K(det(\mathbf{H})^{-1}(\mathbf{x} - \mathbf{x_{ij}}))\, y_{ij}}{\sum_{j=1}^{n_i} K(det(\mathbf{H})^{-1}(\mathbf{x} - \mathbf{x_{ij}}))} \tag{1.9}$$

A complete discussion about the multivariate local regression was shown in Wand and Jones (1995) and Hardle et al. (2004). If the null hypothesis is valid, one could use the

total sample to estimate the common regression; that is,

$$\tilde{g}_{\mathbf{H}}(\mathbf{x}) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K(det(\mathbf{H})^{-1}(\mathbf{x}-\mathbf{x_{ij}}))y_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K(det(\mathbf{H})^{-1}(\mathbf{x}-\mathbf{x_{ij}}))} \tag{1.10}$$

For simplicity, bandwidth **H** was chosen to be equal for both sample specific and total sample kernel functions. The test statistic T_1' for surface comparison proposed by Bowman was

$$T_1' = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\tilde{g}_{\mathbf{H}}(\mathbf{x}) - \tilde{g}_{\mathbf{H}i}(\mathbf{x})]^2}{\hat{\sigma}_B^2} \tag{1.11}$$

where $\hat{\sigma}_B^2 = \frac{1}{2(N-k)}\sum_{i=1}^{k}\sum_{j=1}^{n_i-1}(y_{ij-1} - y_{ij})^2$ from Bock et al. (2007). The argument that T_B' was a ratio of quadratic forms resembles those in the curve comparison in Section 1.1.1. Similarly, matching moments with shifted and scaled χ^2 distribution was used for p-value calculation.

The accuracy of Bowman's testing method depends on the assumption of equal and normal distribution of variance among groups. In cases where normality does not hold, Bowman (2006) recommended a bootstrap procedure for the calculation of p.

Wang's extension of three test statistics from Dette and Neumeyer

Wang et al. (2010) extended Dette and Neumeyer's first nonlinear curve testing method into surface comparison, which tests the difference between linear combined variance functions in the individual samples and in the combined sample. The kernel function and estimated regression ith sample $\tilde{g}_{\mathbf{H}i}(\mathbf{x})$ and common regression $\tilde{g}_{\mathbf{H}}(\mathbf{x})$ were defined in the same way as those in Bowman's method equation (1.9), (1.10). The variance estimator for the ith sample was defined as $\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{i=1}^{n_i}(y_{ij} - \tilde{g}_{\mathbf{H}i}(\mathbf{x}))^2$; correspondingly,

the variance estimator for the total sample size by assuming a common regression function was $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{k}\sum_{i=1}^{n_i}(y_{ij} - \tilde{g}_{\mathbf{H}}(\mathbf{x}))^2$. It follows that the first test statistic by the variance estimator method is the same as (1.4), denoted by T_2' to distinguish from the univariate case. Due to the slow convergence to normal distribution, p-value is calculated based on the distribution of test statistics under H_0 using a wild bootstrap procedure for finite samples.

The second ANOVA-type statistic proposed by Dette (2001) in (1.5), similar to Bowman (2006), was extended by Wang et al. to surface comparison as

$$T_3' = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\tilde{g}_{\mathbf{H}}(\mathbf{x}) - \tilde{g}_{\mathbf{H}i}(\mathbf{x})]^2. \tag{1.12}$$

Similarly, the asymptotic normality of T_3' has been proved, but a wild bootstrap procedure was implemented for testing in practice.

The third test of the hypothesis for common surface is obtained from the summation of weighted differences between the estimates of individual regression functions, which is also a pairwise comparison of regression surfaces extended from equation (1.6)

$$T_4' = \sum_{1\le i<m\le k}\int [\tilde{g}_{\mathbf{H}i}(\mathbf{x}) - \tilde{g}_{\mathbf{H}m}(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x}$$

where $w(.)$ is a positive weight function. $\hat{g}_{\mathbf{H}i}(\mathbf{x})$ and $\hat{g}_{\mathbf{H}m}(\mathbf{x})(1 \le i < m \le k)$ denote the local smooth estimators for the ith and mth group data. In R package 'fANCOVA', Wang

et al. chose $w(\mathbf{x})$ equal to $f(\mathbf{x})$, and used the empirical version of the above statistic as

$$T_4' = \frac{1}{n} \sum_{1 \le i < m \le k} \sum_{j=1}^{n} [\tilde{g}_{\mathbf{H}i}(\mathbf{x}) - \tilde{g}_{\mathbf{H}m}(\mathbf{x})]^2$$

by rescaling the design matrix of $\mathbf{x}$ to have $n_1 = \cdots = n_k = n$.

In their simulation studies, T_4' was shown to outperform T_1', T_2', and T_3' in most cases, as a much better power under the alternative and a satisfactory Type I error control.

1.1.3 Longitudinal or Clustered Data

In Section 1.1.1, the proposed scaled chi-square test for testing equality of two nonparametric curves by fitting a semiparametric mixed model in equation (1.7) obviously would work for correlated data. Zhang and Lin's simulation results showed that their test had a good Type I error control and enough power when the functions of two groups differ with a relatively large effect size. However, so that the scaled chi-square test would work, the simulation was based on the two groups having the same values of xs, which canceled the biases in the smoothing spline estimate under H_0. When two groups do not have the same values of x, the bias cannot be canceled.

Two other existing publications discussed nonlinear curve or surface comparisons by using correlated data. One is a naive method proposed by Bowman, who adopted a simple ad-hoc approach to estimate the pooled random effect and independent measurement variance from the residuals based on the fitted nonparametric surfaces (Bowman, 2006). It is noticed that the bias inherent in smoothing due to the correlation is likely to inflate the variance of these residuals. In Bowman's paper, the bias was regarded

as a conservative effect for comparing curve or surface differences. Let Y_{ijl} be the ith group subject j at lth visit, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n_i$, and $l = 1,2,\ldots,n_i$, where k is the number of groups, n_i is the number of subjects in each group, and n_i is the follow-up visits for the jth individual. The statistical model is formalized as

$$Y_{ijl} = g_i(x_{1ijl}, x_{2ijl}) + \delta_{ij} + \epsilon_{ijl}$$

where (x_{1ijl}, x_{2ijl}) denote explanatory variables, $g_i(x_{1ijl}, x_{2ijl})$ is the nonparametric function of group i, δ_{ij} is the random effects following $N(0,\sigma'^2)$, and $\epsilon_{ijl} \sim N(0,\sigma^2)$. AIC can be used for selecting the smoothing parameter using the same estimators discussed in Section 1.1.1. Due to lack of a simulation study, the validity of testing for this ad-hoc approach was not provided. In addition, the author assumed the same variances of random effect and independent measurement errors for both groups, which may have limited its application in real data analyses.

Wang and Ye (2010) described an indirect bootstrap method for nonparametric surface comparison with spatial correlated errors. First, they suggested estimating the spatial correlation using Francisco-Fernandez and Opsomer's method (Francisco-Fernandez and Opsomer, 2005). In their application, a suitable model is constructed as

$$Y_{ijl} = g_i(x_{1ijl}, x_{2ijl}) + \eta_{ijl}$$

An exponential model was adopted as the correlation function, i.e. $Cov(\eta_{ijp}, \eta_{ijq}) = \sigma_i^2 exp(-\alpha_i ||\mathbf{x}_{ijp} - \mathbf{x}_{ijq}||)$. They estimated the correlation model parameter (σ_i^2, α_i) us-

ing an empirical semivariogram approach. A simple estimator, $\hat{\sigma}^2$, is first calculated based on an average of squared residuals from a pilot fit using a local linear regression kernel based method and a pilot bandwidth matrix. The estimators for α are further derived from the empirical semivariogram. Francisco-Fernandez and Opsomer proved that when the residuals were obtained from a pilot fit, under certain regularity assumptions and an assumption that the correlation coefficient vanishes as the distance goes to infinity with the vanishing speed not slower than $O(1/n)$, both $\hat{\sigma}^2$ and $\hat{\alpha}$ were consistent estimators. Second, "whitened" bootstrap residuals can be generated by using the estimated covariance matrix. Following that, new responses are defined by combining the estimated regression function from the overall observations and "whitened" bootstrap residuals. In the end, the distribution of the test statistic under the null is estimated by the empirical distribution of the statistics generated from the bootstrap samples. However, in their simulation study, their approach did not show a satisfactory power when comparing different surfaces. Wang and Ye explained that this result was due to a large bias in estimating the regression surface with spatial correlated errors.

There are two major limitations of this method relevant to the exponential model assumption. On one hand, exponential correlation decays too fast as the distance grows - an exponential model essentially means that the correlation is dominated by the values near to the origin point. On the other hand, correlations rarely vanish to zero, even at further distance or after a long period of time, where subjects enrolled in longitudinal studies can be a good counter example.

Table 1.1 summarizes the available tests for comparison.

Author(s), Methods	Same x(s)	Correlation	>2 Groups	Curve/Surface	Additional Comments
Fan (1998): Adaptive Neyman test, wavelet thresholding	Y	N	N	Curve	Sensitive to detect local features (e.g. Spectral density analyses, densitometric tracings)
Yong & Bowman (2006): χ^2 approx. w. kernel-based estimators	Y	N	N	Curve/Surface	(+) Simple to implement and understand as a derivation from ANOVA test; (-) Assume equal variance across groups
Dette & Neumeyer (1999), Wang & Ye (2010): Three test statistics	N	N	Y	Curve/Surface	(+) Demonstrated asymptotic normality of all three statistics under H_0; Recommended wild bootstrap when studying finite samples
Zhang & Lin (1998): χ^2 approx. w. semipara. additive	Y	Y	N	Curve	(-) Biased with different values of explanatory variables
Wang & Ye (2010): Spatial correlation	N	Y	Y	Curve/Surface	(-) Larger bias in estimating regression surface hence decreased power

Table 1.1: Summary of the existing methods

Chapter 2

A bootstrap test for curves in semiparametric regression analysis

2.1 Proposed Method: Testing Statistic Based on Semiparametric Regression Estimation with Cross-sectional Data

2.1.1 Review of spline bases and semiparametric regression

Semiparametric regression has been used with an increasing frequency in real data applications, for many good reasons: (1) The approach offers combined advantages of parametric and nonparametric regression models: the former provides a familiar modeling structure and inference for the nonlinear effects, while the latter adds an extra flexibility in accommodating nonlinear effects through the use of low-rank penalized splines. (2) Semiparametric regression models are generally easy to implement. Because they can be written in the form of mixed effect models, they can be fitted using the traditional mixed effect model fitting procedures, in common computing platforms such as SAS and R (Ruppert et. al., 2003).

The testing of linear effects in a semiparametric regression can be carried out as in parametric models. Testing of the nonlinear effects amounts to comparison of nonlinear curves and surfaces. One could, of course, use one of the previously described methods to compare the nonparametric components in the semiparametric models, but that would create a peculiar situation where the estimation and inference of the nonlinear functions are done separately, possibly using different smoothing techniques. For an-

alysts familiar with the penalized spline-based semiparametric regression models, this would lead to a conceptually confusing situation with increased analytical burden.

To alleviate this issue, we propose an inference procedure for the comparison of curves and surfaces, in the usual context of semiparametric regression models. Let Y_{ij} be the subject j in ith group, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n_i$, where k is the number of groups. n_i is the number of subjects in each group. For simplicity, we will discusse here the situations where Y_{ij} is a continuous variable. Without loss of generality, we write the target of the inference as bivariate functions $g_i(x_{1ij}, x_{2ij})$, which represents the nonlinear surfaces associated with the ith group. We note that the approach can easily be extended to comparisons of higher dimensional functions. Therefore, we write the model as $Y_{ij} = g_i(x_{1ij}, x_{2ij}) + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2)$. Function g can be estimated with any smoothing techniques; for example, one expresses g_i, a univariate function for group i, as $g_i(x) = \sum_{k=1}^{K_n+m} \gamma_k B_k^m(x)$, where K_n is the number of internal knots with $K_n = O(n^v)$, m is the order of the B-spline with $m \geq 1$, $\{\gamma_k\}_{k=1}^{K_n+m}$ is the set of the unknown coefficients or control points for the B-spline, and $\{B_k^m(x) : x \in [a,b]\}$ are the basis functions. (Detailed proof in Chapter 3) For multivariate functions $g(x_1, x_2, \ldots, x_d)$, one could choose to use radial or thin-plate splines (Ruppert et al., 2009).

Semiparametric regression, being a well developed regression method, can incorporate various basis functions, automatic or arbitrary number of knots, both univariate and multivariate smoothing functions, and different ranks of smootheing. Established software such as SAS and R are available to implement semiparametric regression either directly (PROC TPSPLINE in SAS and `mgcv`,`gamm4` packages in R) or through mixed model procedures. In the following, we review some commonly used univariate and multivariate semiparametric regression methods.

Univariate spline models

Many types of basis functions have been proposed for univariate smoothing, such as B-splines, thin plate regression splines, and cubic regression splines. For illustration and proof here, we chose B-splines, which is a generation of Bezier curve. Basic references of B-splines can be found in de Boor (1978). Given a vector known as the knot vector defined as $\mathbf{X} = x_0, x_1, \ldots, x_{K_n+1}$, where $\mathbf{X}$ is a nondecreasing sequence and defines control points α_i with $i = 1, \ldots, n$ and n is the number of the basis function. Let $B_i(x)$ denote the value at x of the ith B-spline of order m (degree $m-1$, $n = K_n + m$), then the B-spline curve is

$$B(x) = \sum_{j=1}^{n} \alpha_j B_j(x; p)$$

The general properties of a B-spline of order m include six components: (1) it consists of m polynomial pieces, each of degree $m+1$; (2) the polynomial pieces join at $m+1$ internal knots; (3) At the joining points, derivatives up to order $m-2$ are continuous; (4) the B-spline is positive on a domain spanned by $m+1$ knots, while anywhere else is zero; (5) except at the boundaries, it overlaps with $2(m-1)$ polynomial pieces of its neighbors; and (6) at a given x, m B-splines are nonzero as noted in Eilers and Marx (1996). De Boor provided an algorithm for computing polynomials of any degree from B-splines.

Figure 2.1 shows B-spline bases of degrees one, two, and three for the cases of four irregularly spaced knots. The four irregularly spaced internal knots are given by (0.1, 0.4, 0.6, 0.75). In our further simulation studies, we choose to use a B-spline basis function of degree three, namely cubic B-spline. In this cubic B-spline, the number of

basis functions is $n = K_n + (m-1) - 1 = 4 + 3 - 1 = 6$. The six spline basis functions can be denoted $B_1, \ldots, B_6$ following the above notations.

B-spline basis is very useful in practice, because it avoids the disadvantage of polynomial or truncated power bases that the explanatory variables as being far from orthogonal, which lead to numerical instability. Hence, Ruppert et al. (2003) stated that for ordinary least square fitting, the most common choice is B-spline bases.

For model construction with cubic B-spline bases, if x_{ij} is the non-linear explanatory variable for prediction, and the continuous response variable Y_{ij}, the cubic B-spline smoothing function $g_i(.)$ for group i in the univariate semiparametric model

$$Y_{ij} = g_i(x_{ij}) + \epsilon_{ij} \tag{2.1}$$

can be expressed as $g_i(x) = \sum_{j=1}^{n_i} \alpha_{ij} B_j(x; p)$, where α_{ij} are the unknown control points and p is three.

Univariate penalized spline models for a smooth real-valued function g spline can be expressed using a mixed model-based penalized spline approach

$$g(x; p, \mathbf{z}) = \beta_0 + \cdots + \beta_p x^p + \sum_{k=1}^{K} \mu_k z_k(x)$$

where p is the degree of polynomial component with coefficients $\beta_0, \ldots, \beta_p$ and $\mathbf{z}$ is a set of spline basis functions. A simple example is $z_k(x) = (x - k_k)_+$ for some knot sequences $k_1, \ldots, k_K$. The spline coefficients $\mu = (\mu_1, \ldots, \mu_K)$ are subject to penalization in Ruppert et al. (2009). Most of the spline bases are in accordance with the classical nonparametric regression method known as smoothing splines, which includes thin plate regression

splines, and cubic regression splines. Cubic B-spline is commonly used with degree of the piecewise polynomial as three.

Multivariate spline models

The smoothing technique of univariate spline models has been extended to the multivariate case using either a radial basis function or tensor products. Ruppert et al. (2003) summarized bivariate smoothing approaches based on kriging and splines. Wood developed low-rank thin plate spline smoothing (Wood, 2003) and tensor products (Wood, 2006). We focus on the illustration of two of Wood's smoothing methods, as they are more commonly used and easy to implement in practice.

The low-rank thin plate spline smoothers have been constructed by a transformation and truncation of the basis and are optimal in the sense that the truncation is designed to result in the minimum possible perturbation of the thin plate spline smoothing given the dimension of the basis. This circumvents the knot placement issue. We consider a problem of estimating the smooth function of two predictors $g(x_1, x_2)$ for n observations such that

$$y = g(x_1, x_2) + \epsilon \tag{2.2}$$

where ϵ is a random error term. g is estimated by finding a function $\hat{g}$ that minimizes $||\mathbf{y} - \mathbf{g}||^2 + \lambda J_m(g)$, where $\mathbf{g} = (g(\mathbf{x}_1), g(\mathbf{x}_2), \ldots, g(\mathbf{x}_n))'$, $||.||$ is the Euclidean norm, q is the order of differentiation in the penalty term, and $J_q(g)$ is a penalty function. If wiggliness

is measured using second derivatives, $J_2(g)$ is

$$J_2(g) = \int\int (\frac{\partial^2 g}{\partial x_1^2})^2 + 2(\frac{\partial^2 g}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 g}{\partial x_2^2})^2 dx_1 dx_2$$

Solving function g involves a computational burden of high rank matrix calculation when finding the resulting smoothing objective. However, when seeking the ideal smooth, low rank smoothers can be constructed. The truncation of high rank matrices into lower rank matrices is obtained by minimizing the 'worst' possible (maximum) change of a weighted difference in a form of Euclidean norm between the original high rank matrix and the truncated matrix. For small datasets, thin plate regression spline can be implemented using routine linear algebra; but for large datasets, it is necessary to obtain thin plate regression spline bases using Lanczos iteration.

A noteworthy feature of the thin plate spline approach is the isotropy of the wiggliness penalty, i.e. wiggliness in all directions is treated equally. However, one major disadvantage of this isotropy is that it is difficult to know how to scale predictors relative to one another, when both are arguments of the same smooth but measure in different units. This method starts from smoothing single covariates, followed by constructing a 'tensor product' of several variables with these 'marginal smooths'. Wood (2006) considered the same model in (2.2). Firstly, two marginal bases and penalties are obtained, as if two univariate smooth terms $g_{x_1}(x_1)$ and $g_{x_2}(x_2)$. Let the bases be $a_{x_1 i}(x_1) : i = 1, \ldots, I$ and $a_{x_2 k}(x_2) : k = 1, \ldots, K$ with associated parameters α_i and β_k. i.e. $g_{x_1}(x_1) = \sum_{i=1}^{I} \alpha_i a_{x_1 i}(x_1), g_{x_2}(x_2) = \sum_{k=1}^{K} \beta_k a_{x_2 k}(x_2)$. In order to convert the smooth function of x_1 into a smooth function of x_1 and x_2, g_{x_1} are required to vary smoothly with x_2, which can be achieved by allowing the parameters α_i to vary smoothly with x_2,

i.e. $\alpha_i(x_2) = \sum_{k=1}^{K} \beta_{ik} a_{x_2 k}(x_2)$. By the usual tensor product construction, the smooth g is represented as a 'tensor product' basis

$$g(x_1, x_2) = \sum_{i=1}^{I} \sum_{k=1}^{K} a_{x_1 i}(x_1) a_{x_2 k}(x_2) \beta_{ik}$$

where the IK coefficients β_{ik} are unknown. Secondly, it is necessary to have some way of measuring tensor product penalties. Suppose the penalty with coefficient matrix S_{x_1} and S_{x_2} penalize the marginal coefficients associated with the sequence of basis functions $a_{x_1 1}(x_1), a_{x_1 2}(x_1), \ldots, a_{x_1 I}(x_1)$ and $a_{x_2 1}(x_2), a_{x_2 2}(x_2), \ldots, a_{x_2 K}(x_2)$ respectively. i.e. $J_{x_1}(g_{x_1}) = \alpha^T S_{x_1} \alpha, J_{x_2}(g_{x_2}) = \beta^T S_{x_2} \beta$. The $S_\cdot$ matrices contain known coefficients, and α and β are vectors of coefficients of the marginal smooths. An example of a penalty functional is the cubic spline penalty, $J_x(g_x) = \int (\partial^2 g_x / \partial x^2)^2 dx$. Now let $g_{x_1|x_2}(x_1)$ be g_{x_1,x_2} considered as a function of x_1 only, with x_2 held constant, and define $g_{x_2|x_1}$ similarly. A natural way of measuring wiggliness of $g_{x_1 x_2}$ is to use

$$J(g_{x_1 x_2}) = \lambda_{x_1} \int_{x_2} J_{x_1}(g_{x_1|x_2}) dx_2 + \lambda_{x_2} \int_{x_1} J_{x_2}(g_{x_2|x_1}) dx_1$$

where $\lambda_\cdot$ are smoothing parameters controlling the tradeoff between wiggliness in different directions and allowing the penalty to be invariant to the relative scaling of the covariates. If cubic spline penalties are used as the marginal penalties, then $J(g_{x_1 x_2}) = \int_{x_1,x_2} \lambda_{x_1} (\frac{\partial^2 g}{\partial x_1^2})^2 + \lambda_{x_2} (\frac{\partial^2 g}{\partial x_2^2})^2 dx_1 dx_2$. Hence, if the marginal penalties are easily interpretable, then so is the induced penalty. For example, if we considered the penalty in the x_1 direction, the function $g_{x_1|x_2}(x_1)$ can be written as $g_{x_1|x_2}(x_1) = \sum_{i=1}^{I} \alpha_i(x_2) a_i(x_1)$ and it is possible to find the matrix of coefficients M_{x_2} such that $\alpha(x_2) = M_{x_2} \beta$, where β is the

vector of β_{ik} arranged in the appropriate order. Hence $J_{x_1}(g_{x_1|x_2}) = \alpha(x_2)^T S_{x_1} \alpha(x_2) = \beta^T M_{x_2}^T S_{x_1} M_{x_2} \beta$ and so $\int_{x_2} J_{x_1}(g_{x_1|x_2}) dx_2 = \beta^T \int_{x_2} M_{x_2}^T S_{x_1} M_{x_2} dx_2 \beta$. The last integral can be performed numerically, but a simple re-parameterization is used to provide an approximation to the terms in the penalty to avoid explicit numerical integration. In the end, the approximate $J(g_{x_1 x_2})$ can be written as $J(g_{x_1 x_2}) \approx J^*(g_{x_1 x_2}) = \lambda_{x_1} J_{x_1}^*(g_{x_1 x_2}) + \lambda_{x_2} J_{x_2}^*(g_{x_1 x_2})$, where J^* is an approximation after re-parameterization.

2.1.2 The test statistic and a wild bootstrap-based comparison method

An intuitive way to compare two functions is to measure the distance between them. The L_2 norm is one of the most frequently used distance measures for this purpose. We note that both Zhang et al. (2000) and Neumeyer and Dette (2003) both used the L_2 norm in their construction of the test statistics. Herein, we reexamine the test statistic

$$T_{spline} = \frac{1}{N} \sum_{1 \le i < m \le k} \sum_{j=1}^{n_i} [\hat{g}_i(\mathbf{x}_{ij}) - \hat{g}_m(\mathbf{x}_{ij})]^2,$$

under the B-spline estimates of $\hat{g}_i$ and $\hat{g}_m$. We show that under fairly general conditions, the test statistic is consistent.

But in the absence of an asymptotic distribution, we have to devise a method through which we can approximate the distribution of the test statistic under the null hypothesis. In this section, we demonstrate how such an approximation is done through a resampling procedure. Specifically, we show how p values for the test statistic can be ascertained from a wild bootstrap procedure.

For the standard linear regression model, it is usually sufficient to draw bootstrap samples from the centralized residuals, because the errors are homoscedastic. In the

current research, the underlying model is $E(Y_{ij}|\mathbf{X}) = g_i(X_{ij}) + \epsilon_{ij}(X_{ij})$, where the errors are clearly heteroscedastic. To accommodate the error heteroscedasticity, we consider a wild bootstrap procedure, which assures the bootstrap error terms have properties that are similar to those of the actual errors (Liu, 1988). Another alternative is the pairs bootstrapping, in which the analysts directly resample from the empirical distribution function of Y_i and $\mathbf{X}_i$. The computational burden, in this case, however, is much greater (Freedman, 1981).

For nonparametric models, wild bootstrap has been used to resample the residuals of nonparametric regression models, as done by Hardle and Mammen (1993) and Mammen (1993). The essence of the wild bootstrap is to express the regression function as a conditional expectation of the observed response variable, i.e. $E(Y_i^*|X_i) = g(X_i)$, where Y_i^* is the bootstrap data. Since this method uses a single residual $\hat{\epsilon}_i$ to estimate the conditional distribution $l(Y_i - g(x_i)|X_i = x_i)$ by an arbitrary distribution ($\hat{F}_i$ in the following), it is often referred to as the wild bootstrap.

Let V_i be a random variable following a two-point distribution $\hat{F}_i$ such that $E_{\hat{F}_i}(V_i) = 0, E_{\hat{F}_i}(V_i^2) = 1, E_{\hat{F}_i}(V_i^3) = 1$. We construct independent $\epsilon_i^* = V_i\hat{\epsilon}_i \sim \hat{F}_i$ and use $(X_i, Y_i^* = \hat{g}(x_i) + \epsilon_i^*)$ as the bootstrap observations. We then create a new bootstrap test statistic T^*.

With the bootstrap samples, for a test at level α, the null hypothesis is rejected if T is greater than the corresponding quantile of the bootstrap distribution T^*, i.e. $T > T^*_{(B(1-\alpha))}$, where $T^*_{(B(1-\alpha))}$ is the ith order statistics of the bootstrap statistic T^*. Hardle and Mammen demonstrated that under the null hypothesis, the wild bootstrap T^* estimated the distribution of T consistently, since the regression function with boot-

strap data $g^*(.)$ had mean $g(.)$ for nonlinear models under the standard regularity conditions.

Using the wild bootstrap method, we propose the following testing procedure:

1. Estimate $\hat{g}_i(\mathbf{x})$ from the two groups separately and compute the test statistic $T_{spline} = \frac{1}{N}\sum_{1\le i<m\le k}\sum_{j=1}^{n_i}[\hat{g}_i(\mathbf{x}_{ij}) - \hat{g}_m(\mathbf{x}_{ij})]^2$.
2. Estimate the common regression surface $\hat{g}(\mathbf{x})$ from the combined sample and calculate the residuals $\hat{\epsilon}_{ij} = y_{ij} - \hat{g}(\mathbf{x}_{ij})$.
3. For each $\mathbf{x}_{ij}$, draw a bootstrap residual ϵ^*_{ij} from the two-point distribution with probability masses $\frac{1-\sqrt{5}}{2}\hat{\epsilon}_{ij}$ and $\frac{1+\sqrt{5}}{2}\hat{\epsilon}_{ij}$, occurring with probabilities $\frac{5+\sqrt{5}}{10}$ and $\frac{5-\sqrt{5}}{10}$ respectively, so that $E(\epsilon^*_{ij}) = 0$, $E(\epsilon^{*2}_{ij}) = \hat{\epsilon}^2_{ij}$ and $E(\epsilon^{*3}_{ij}) = \hat{\epsilon}^3_{ij}$.
4. Generate a bootstrap sample $(\mathbf{x}_{ij}, Y^*_{ij})$ by setting $Y^*_{ij} = \hat{g}(\mathbf{x}_{ij}) + \epsilon^*_{ij}$.
5. From this sample, calculate the bootstrap regression surfaces $\hat{g}^*_i$ and the test statistic T^*_{spline} in the same way as the original T_{spline} is calculated.
6. Repeat steps (3) to (5) B times and use the B generated test statistics T^*_{spline} to determine the quantiles of the test statistic under the null hypothesis. For a test at significance level α, the null hypothesis is rejected if T_{spline} is greater than the corresponding quantile of the bootstrap distribution of T^*_{spline}.

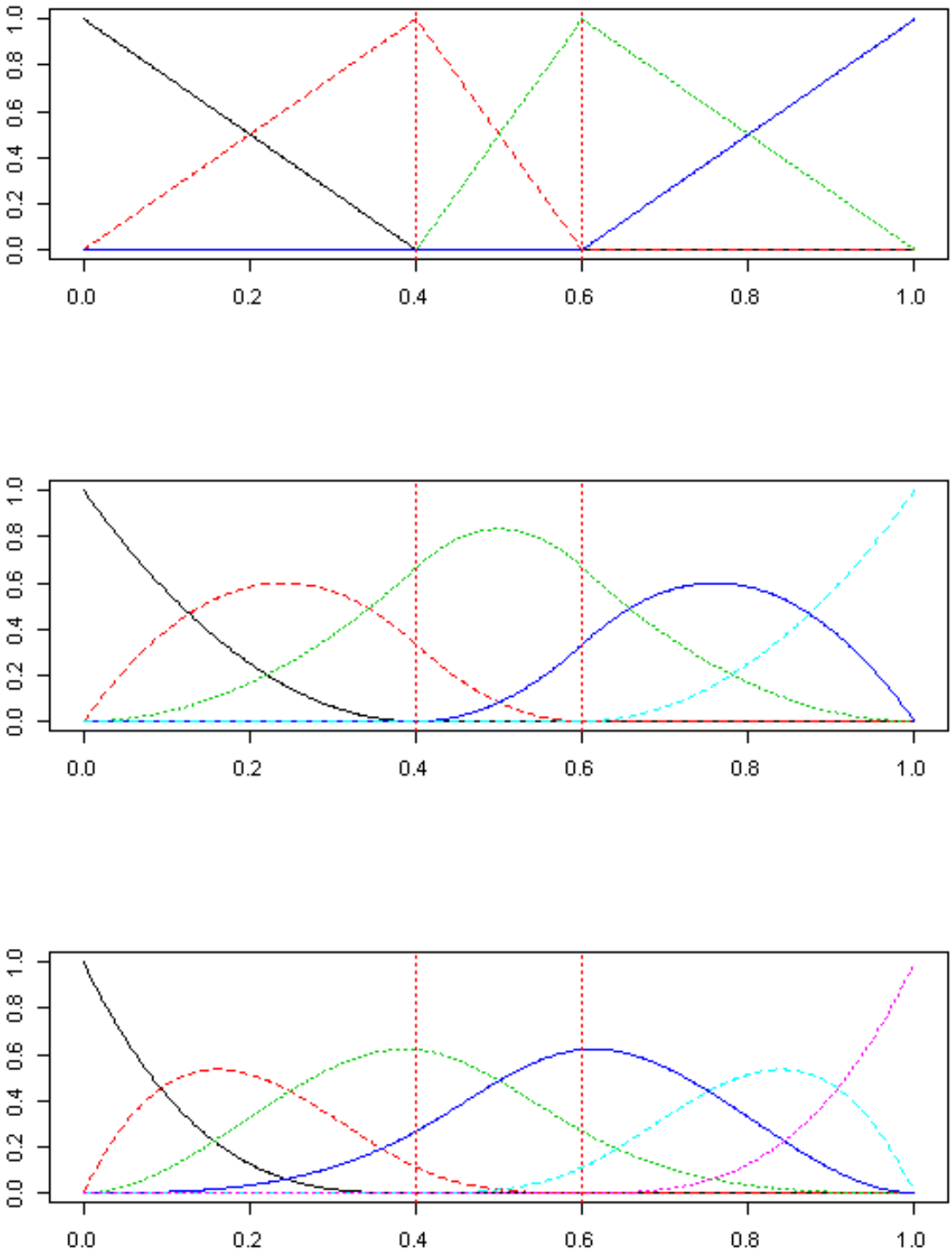

Figure 2.1: B-spline bases of degrees one, two, three

2.2 Simulation Studies

Simulation studies were performed to compare the performance of our testing method with existing kernel based methods for curve comparisons and surface comparisons. Moreover, we investigated the effects of number of knots selection on rejecting probabilities.

2.2.1 Simulation studies for curve comparisons

We considered the following model in the simulation

$$Y_{ij} = g_{id}(x_{ij}) + \epsilon_{ij}. \tag{2.3}$$

where $i = 1,2; j = 1,\ldots,n_i$. Values of independent variable x_{ij} were generated from $Unif[0,1]$ independently, with a sample size of n_1 and n_2 for each group. The nonlinear functions for the two groups were specified as $g_1(X) = 2X\exp(2-4X) - 2X + 0.5$ and $g_2(X) = 2X\exp(2-4X) - 0.5$ with an average between the two $g(X) = (4X\exp(2-4X) - 2X)/2$. More generally, we considered $g_{id}(X_{ij}) = (d/10)g_i(X_{ij}) + (1-d/10)g(X_{ij})(d = 0,1,2,3)$, where d controls the distance between the two group-specific functions. For example, $d = 0$ corresponded to the situation where the two groups shared the same regression function: as d increased, the functions grew further apart. These functions were plotted in Figure 2.2. Values of the dependent variables Y_{ij} were generated from Equation (2.3) with standard error σ_1 and σ_2, i.e. $\epsilon_{1j} \sim N(0,\sigma_1^2)$, $\epsilon_{2j} \sim N(0,\sigma_2^2)$. Data simulation was performed under the control of three sets of input parameters: (1) $d =$

0,1,2,3; (2) the sample sizes (n_1, n_2)=(125, 125), (216, 216), and (512, 512); (3) standard deviation of the errors (σ_1, σ_2)=(0.20, 0.15) and (0.25, 0.20).

Thereby, we compare the performance of five different methods:

- Method 1: The proposed testing method with cubic B-spline regression bases for curve estimation, with a wild bootstrap procedure for p-value calculation. Knots numbers were chosen as $\sqrt[3]{n_1}$.
- Method 2: The proposed testing method with a penalized cubic spline basis for curve estimation, with a wild bootstrap procedure for a p-value calculation by using the `gam` function in R package `mgcv`. Numbers of knots numbers were set to the default value, which was determined by a generalized cross-validation (GCV) method.
- Method 3: Kernel smoothing based on the L^2 distance test statistic, followed by a wild bootstrap, as described by Dette and Neumeyer.
- Method 4: The testing method based on the variance estimator.
- Method 5: Bowman's method, which calculated p value by matching with a scaled chi-square distribution.

For each simulation setting, we generated 1000 testing datasets. For each test, we used 200 wild bootstrap samples for calculation of the p values. We calculated the rejection rate out of the 1000 simulation using a significance level of 0.05.

Table 2.1 summarizes the rejection probabilities for curve comparison under the null and alternative hypotheses. Notations for each testing method in the table below are consistent with those presented in the main manuscript. $T_{B-spline}$: L^2 distance of pointwise B-spline based estimating regression functions with $k = \sqrt[3]{n_i}$; $T_{P-spline}$: P-spline estimating regression function with default number of knots from GCV ($k \approx 6$ in

this example); T_4: Kernel based estimating regression function $\tilde{g}_i(x)$, $T_4 = \sum_{i=1}^{k}\sum_{j=1}^{i-1}[\tilde{g}_i(x) - \tilde{g}_j(x)]^2$; T_2: variance estimating method, $T_2 = \hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{k} n_i\hat{\sigma}_i^2$; T_1: "Bowman's" test matching with a scaled chi-square distribution, $T_1 = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\tilde{g}(x_{ij}) - \tilde{g}_i(x_{ij})]^2$. Type I error rates were provided for situations with $d = 0$; and power was provided when $d = 1,2,3$. As d increased, the power for rejecting the null increased as well. When $d \neq 0$, the rejection probability increased with decreased variances and larger sample sizes. The "Bowman's" test showed slightly higher type I errors and greater power than the others. On the other hand, the T_2 test exhibited tighter type I error control, while having considerably lower power. For the proposed testing method, B-splines and penalized splines provided similar results. As previously discussed, we set the number of knots to $\sqrt[3]{n_i}$. Our early simulation results showed when an unpenalized semiparametric model was used, an incorrect number of knots selection could lead to biased estimated regression functions, and thus much inflated type I error rates. The penalized semiparametric estimating methods were generally more robust to the number of knots selection.

2.2.2 Simulation studies for surface comparisons

Simulation for surface comparisons were carried out in a similar manner. The surface functions were specified as follows:

a. $g_1(\mathbf{X}) = g_2(\mathbf{X}) = \sin(2\pi X_1) + \cos(2\pi X_1)$

b. $g_1(\mathbf{X}) = g_2(\mathbf{X}) = 2X_1^2 + 3X_2^2$

c. $g_1(\mathbf{X}) = g_2(\mathbf{X}) = \exp(-X_1^2 - X_2^2)$

d. $g_1(\mathbf{X}) = \sin(2\pi X_1) + \cos(2\pi X_1) \qquad g_2(\mathbf{X}) = \sin(2\pi X_1) + \cos(2\pi X_2) + X_1$

e. $g_1(\mathbf{X}) = 2X_1^2 + 3X_2^2 \qquad g_2(\mathbf{X}) = 2X_1^2 + 3X_2^2 + \sin(2\pi X_1)$

f. $g_1(\mathbf{X}) = \exp(-X_1^2 - X_2^2) \qquad g_2(\mathbf{X}) = \exp(-X_1^2 - X_2^2) + \sin(2\pi X_1)$

Scenarios **a**-**c** represented situations where the surfaces were the same; scenarios **d**-**f** corresponded to the alternative hypothesis. Contour plots and 3-D plots are shown in Figure 2.3, Figure 2.4, and Figure 2.5. Interactive plots are available online at: Scenarios **d** `https://zhaoshi169.github.io/simsurface1.html`; Scenarios **e** `https://zhaoshi169.github.io/simsurface2.html`, and Scenarios **f** `https://zhaoshi169.github.io/simsurface3.html`. The independent variables x_1 and x_2 were simulated from independent $Unif[0,1]$ with a sample size n_1 and n_2 for each group. The dependent variables Y_{ij} were generated from the above functions with a standard error of σ_1 and σ_2, i.e. $Y_{1j} = g_1(X_{1j}) + \epsilon_{1j}$, $Y_{i2} = g_2(X_{2j}) + \epsilon_{2j}$, where $i = 1,2; j = 1,\ldots,n_i, \epsilon_{1j} \sim N(0,\sigma_1^2), \epsilon_{2j} \sim N(0,\sigma_2^2)$.

Simulation was then performed under the following parameter settings: (1) three sample size settings of (n_1, n_2) as (125, 125), and (216, 216), (512,512); (2) Two different values the the standard errors (σ_1, σ_2) for each function.

For each simulation setting, we generated 500 datasets. For each dataset, we tested the new method with 300 wild bootstrap resamples. We calculated the rejection rate based on the 500 simulated datasets with the significance level set at 0.05.

Again, the five statistical methods were compared on each set of two surfaces.

- Method 1: The proposed method of fitting a multivariate semiparametric model for surface estimations then implementing a wild bootstrap procedure for p-value calculation. For the thin-plate based estimatiang method, three knots number selection were examined, including $\sqrt[3]{n_i}$, $\sqrt[3]{n_i}+1$, $\sqrt[3]{n_i}-1$ for x_1 and x_2. Congruently knots number $\sqrt[3]{n_i}$ was selected for the tensor-product estimatiang method. Penalized multivariate semiparametric estimating regression functions were compared with both thin-plate and tensor-product estimating methods with a default number of knots in `gam` function in R on both predictors.
- Method 2: Fitting a penalized semiparametric thin-plate spline or tensor-product regression model for surface estimations and then a wild bootstrap procedure for p-value calculation. The knots numbers used default `gam` function in R package `mgcv` (default number of knots estimated by generalized cross-validation (GCV) method) .
- Method 3: Kernel-smoothing based on the L^2 distance test statistic following by a wild bootstrap for surface hypothesis testing, proposed by Wang (2011).

- Method 4: Bowman's and Dette's ANOVA-type method (test statistics based on 1.11 and 1.12), which calculated the p value by matching with a scaled chi-square distribution.
- Method 5: The testing method based on the variance estimator.

Simulation results for surface comparisons are presented in Tables 2.2 and 2.3. In the table, "TP-Spline", "TP-Spline$_+$" and "TP-Spline$_-$" indicate tests using thin-plate splines with $\sqrt[3]{n_i}$, $\sqrt[3]{n_i}+1$, and $\sqrt[3]{n_i}-1$ knots. "TE" indicates a testing method using tensor-product basis functions, while"TP-Spline.p" and "TE-Spline.p" are tests using penalized splines.

Type I error control was compared by rejection probability across the four methods in function pairs of **a** - **c** and power comparisons by function pairs of **d** - **f**. Coinciding with the simulation studies for curve testings, overall, the penalized semiparametirc model using a default number of knots from 'GCV' method showed a comparable performance with the tests using the semiparametirc estimating methods and an adjusted number of knots. Selection of number of knots had slight influences on the testing significances. In comparing our methods to the previous ones: 1) the semiparametric spline estimation based and nonparametric estimation based L^2 distance testings exhibited similar testing significances; 2) Dette's ANOVA-type method lacked a Type I error control; 3) in contrast, Bowman's ANOVA method and variance estimating approach were over controlled; and 4) our proposed test with penalized or unpenalized spline models exhibited comparable or superior power for rejecting the null hypothesis than the other methods, shown in Table 2.3.

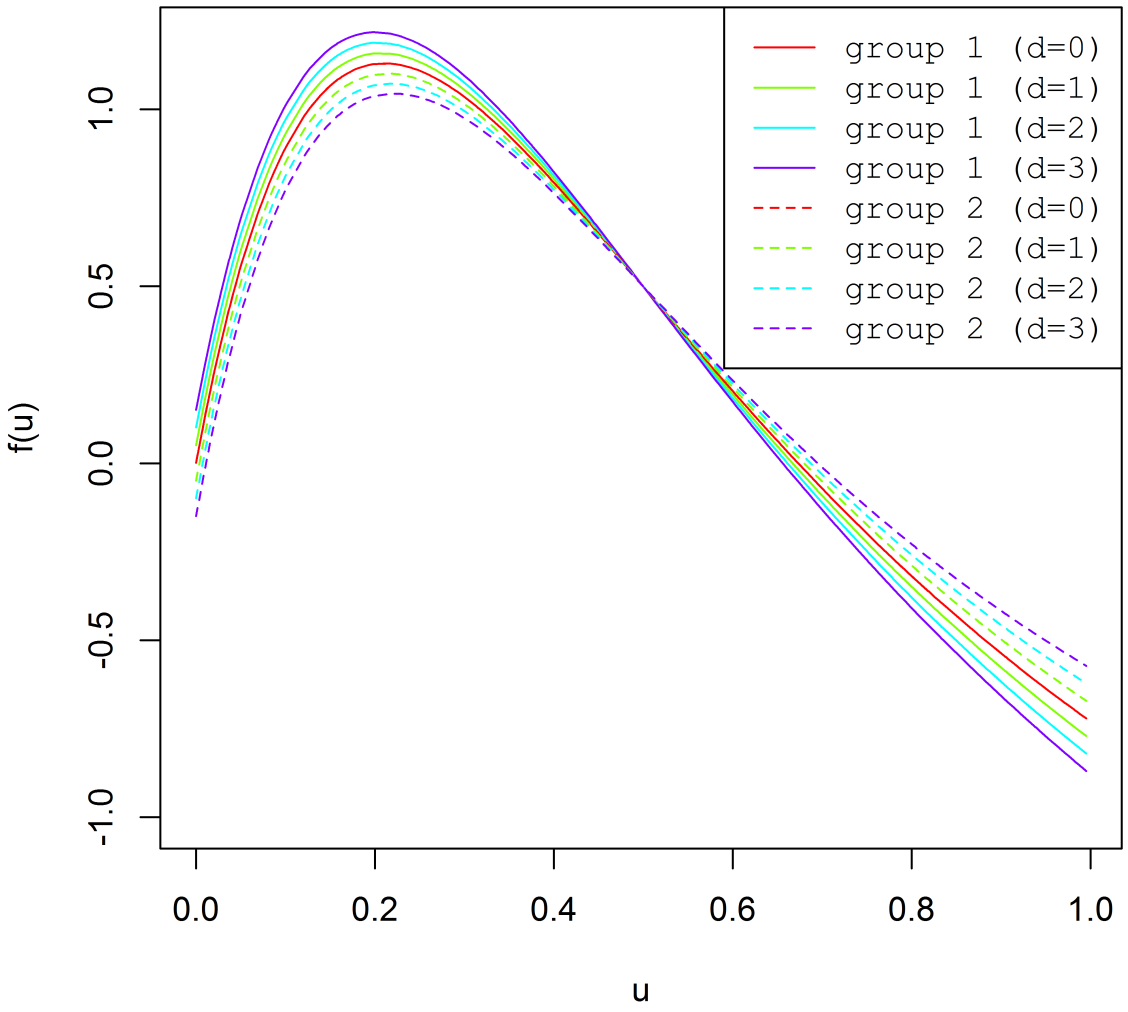

Figure 2.2: Functions $g_{1d}(x)$ and $g_{2d}(x)$ with $d = 0, 1, 2, 3$ used in the simulation studies

d	(n_1, n_2)	(σ_1, σ_2)	$T_{B-spline}$	$T_{P-spline}$	T_4	T_2	T_1
0	(125, 125)	(0.2, 0.15)	0.051	0.049	0.070	0.043	0.060
	(216, 216)	(0.2, 0.15)	0.048	0.057	0.071	0.046	0.067
	(512, 512)	(0.2, 0.15)	0.055	0.060	0.066	0.049	0.063
	(125, 125)	(0.25, 0.2)	0.049	0.051	0.065	0.043	0.060
	(216, 216)	(0.25, 0.2)	0.051	0.052	0.071	0.058	0.072
	(512, 512)	(0.25, 0.2)	0.060	0.047	0.065	0.057	0.066
1	(125, 125)	(0.2, 0.15)	0.416	0.349	0.406	0.309	0.415
	(216, 216)	(0.2, 0.15)	0.688	0.621	0.655	0.529	0.666
	(512, 512)	(0.2, 0.15)	0.969	0.967	0.973	0.926	0.972
	(125, 125)	(0.25, 0.2)	0.274	0.226	0.321	0.208	0.302
	(216, 216)	(0.25, 0.2)	0.434	0.379	0.446	0.353	0.469
	(512, 512)	(0.25, 0.2)	0.824	0.802	0.850	0.735	0.848
2	(125, 125)	(0.2, 0.15)	0.974	0.941	0.956	0.920	0.958
	(216, 216)	(0.2, 0.15)	1.000	1.000	1.000	0.995	1.000
	(512, 512)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.25, 0.2)	0.844	0.764	0.830	0.725	0.821
	(216, 216)	(0.25, 0.2)	0.985	0.970	0.982	0.952	0.983
	(512, 512)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000
3	(125, 125)	(0.2, 0.15)	1.000	1.000	0.999	1.000	0.999
	(216, 216)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(512, 512)	(0.2, 0.15)	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.25, 0.2)	0.995	0.990	0.992	0.981	0.995
	(216, 216)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000
	(512, 512)	(0.25, 0.2)	1.000	1.000	1.000	1.000	1.000

Table 2.1: Curve comparison: Power and Type 1 error rates.

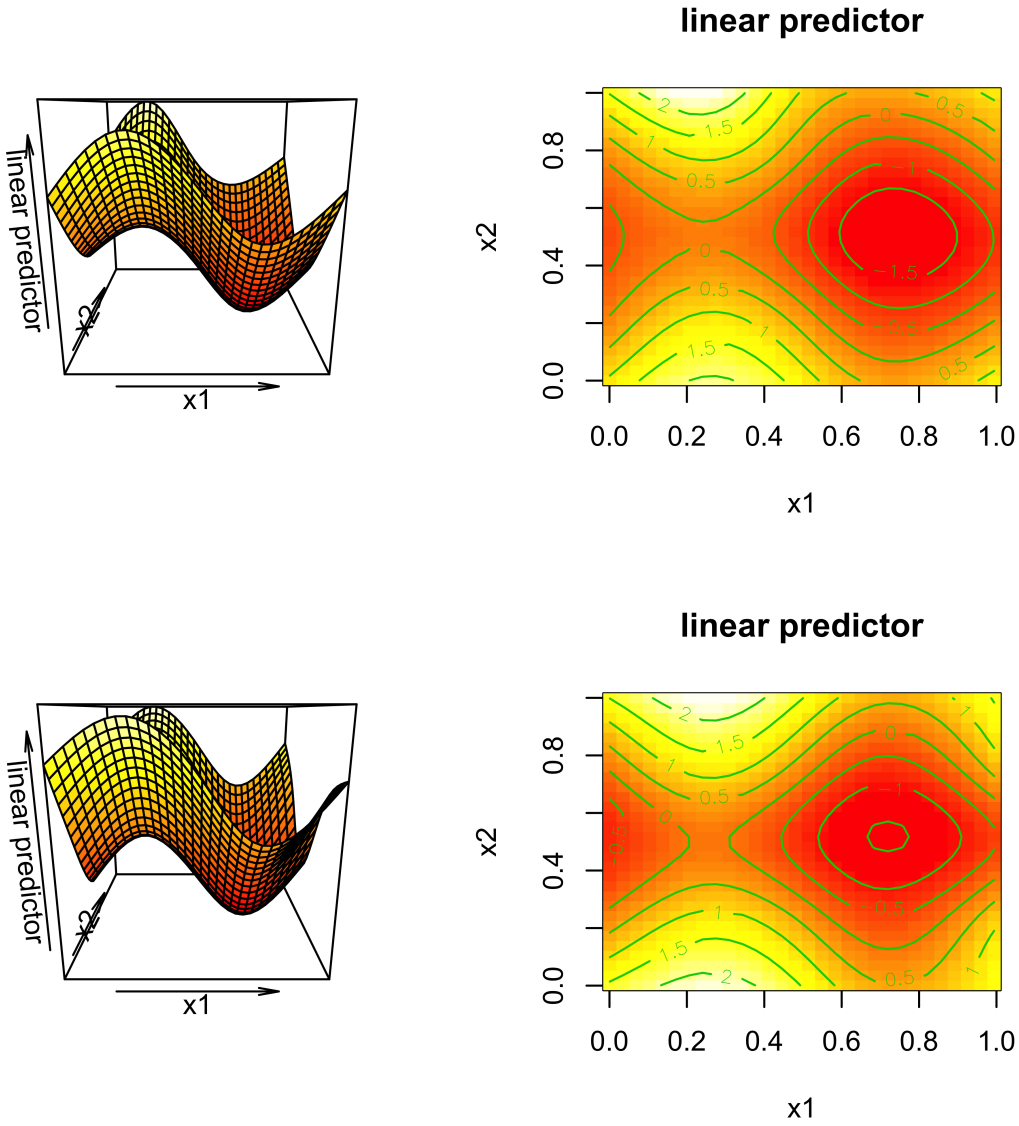

Figure 2.3: Surface functions c used in the simulation studies

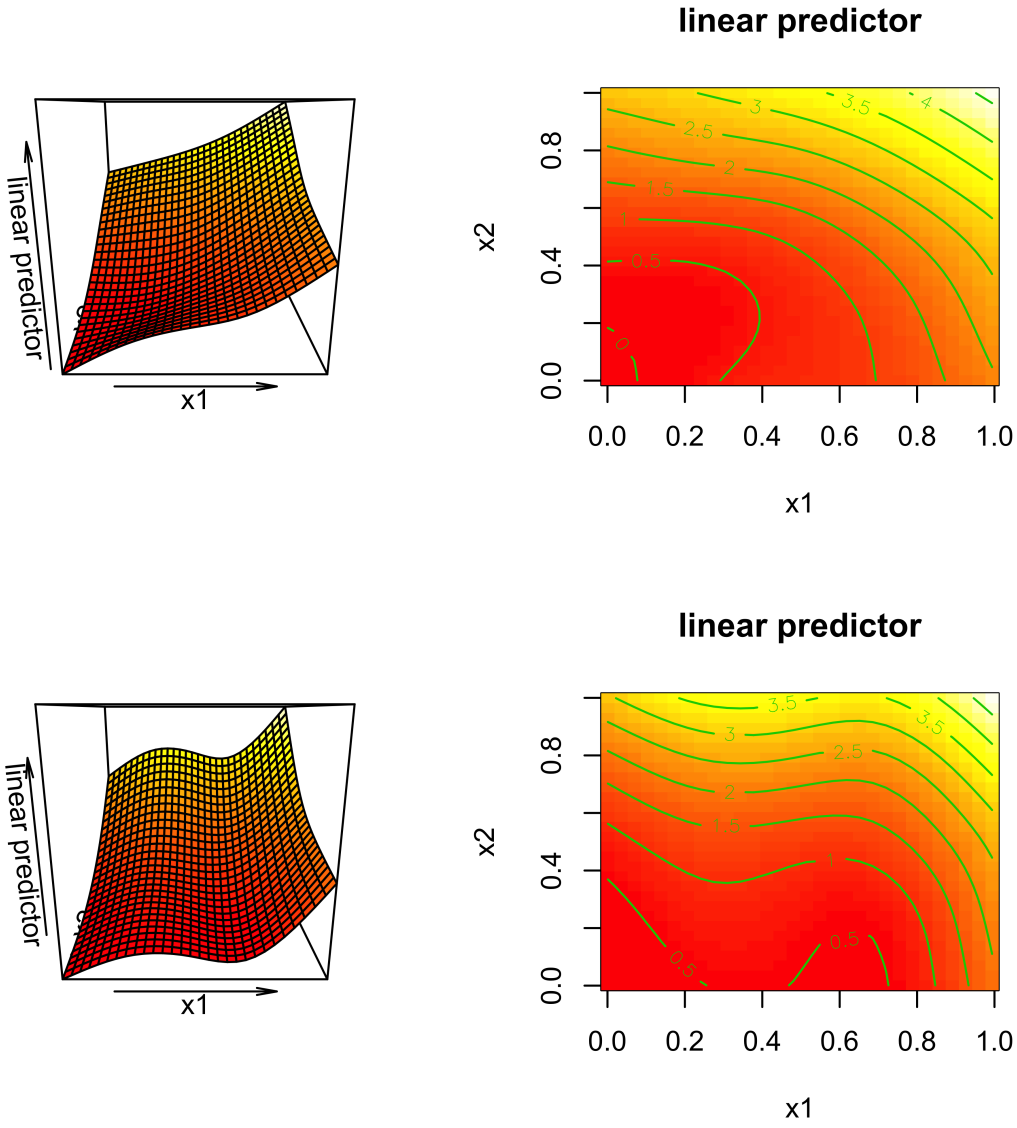

Figure 2.4: Surface functions d used in the simulation studies

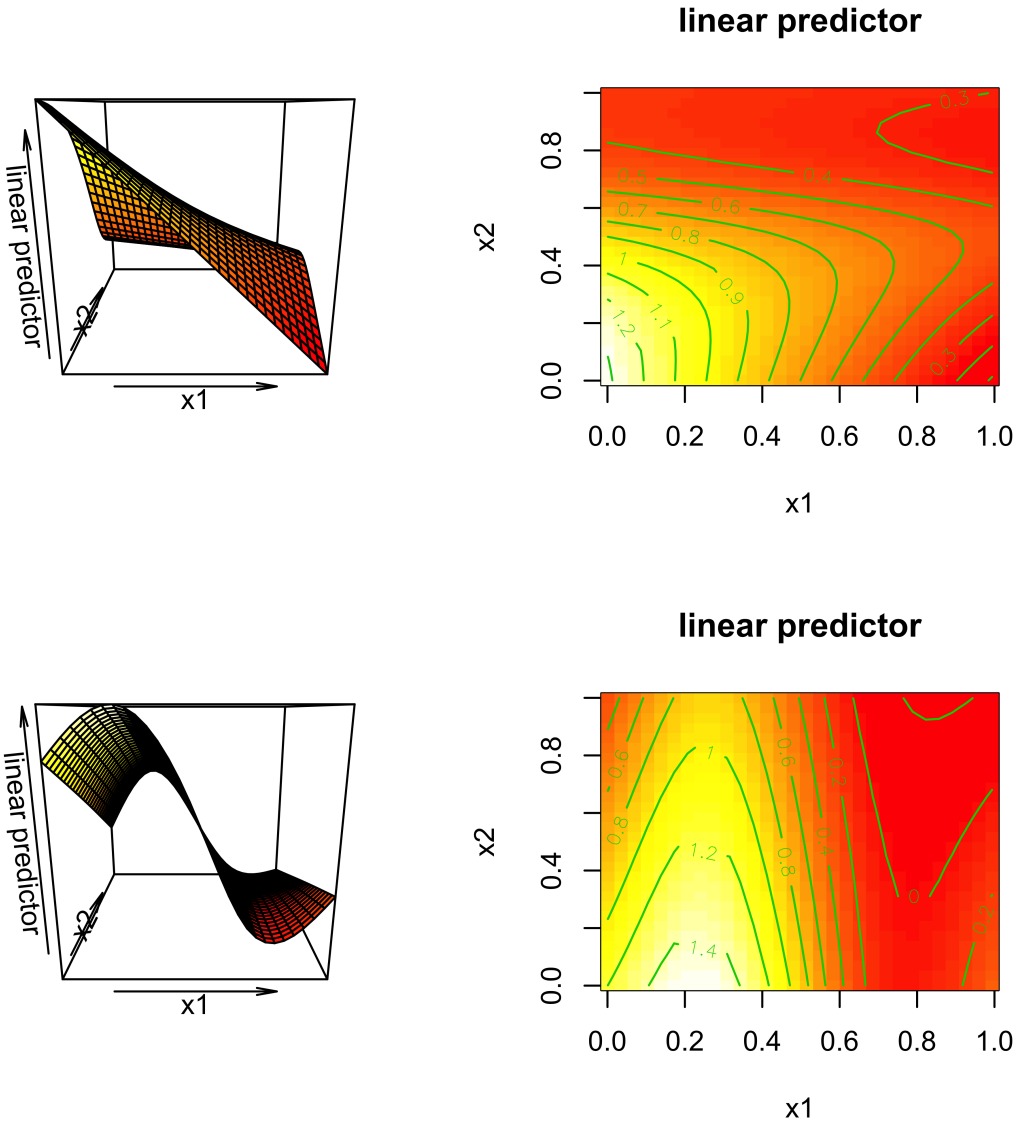

Figure 2.5: Surface functions e used in the simulation studies

Func	(n_1, n_2)	(σ_1, σ_2)	TP-Spline	TP-Spline$_+$	TP-Spline$_-$	TE-Spline	TP-Spline.p	TE-Spline.p	T_4	T_1	T_3	T_2
a	(125, 125)	(0.5, 0.3)	0.088	0.10	0.088	0.10	0.106	0.094	0.09	0.038	0.15	0.002
	(216, 216)	(0.5, 0.3)	0.076	0.080	0.094	0.07	0.07	0.068	0.088	0.02	0.110	0.012
	(512, 512)	(0.5, 0.3)	0.074	0.068	0.058	0.080	0.046	0.06	0.064	0.026	0.064	0.012
	(125, 125)	(0.6, 0.4)	0.098	0.086	0.09	0.070	0.080	0.080	0.08	0.02	0.114	0.002
	(216, 216)	(0.6, 0.4)	0.078	0.080	0.100	0.07	0.078	0.068	0.078	0.02	0.120	0.010
	(512, 512)	(0.6, 0.4)	0.05	0.06	0.07	0.060	0.06	0.056	0.056	0.016	0.08	0.012
b	(125, 125)	(0.6, 0.4)	0.06	0.05	0.044	0.068	0.048	0.060	0.058	0.066	0.060	0.052
	(216, 216)	(0.6, 0.4)	0.044	0.038	0.038	0.034	0.044	0.038	0.070	0.080	0.088	0.026
	(512, 512)	(0.6, 0.4)	0.054	0.048	0.05	0.040	0.05	0.05	0.080	0.07	0.076	0.038
	(125, 125)	(0.8, 0.6)	0.060	0.058	0.06	0.060	0.05	0.064	0.074	0.080	0.078	0.068
	(216, 216)	(0.8, 0.6)	0.066	0.07	0.064	0.066	0.05	0.05	0.074	0.076	0.084	0.038
	(512, 512)	(0.8, 0.6)	0.064	0.058	0.06	0.050	0.058	0.050	0.060	0.060	0.06	0.026
c	(125, 125)	(0.8, 0.6)	0.038	0.040	0.04	0.056	0.048	0.040	0.040	0.046	0.04	0.056
	(216, 216)	(0.8, 0.6)	0.05	0.05	0.054	0.048	0.046	0.03	0.046	0.050	0.048	0.056
	(512, 512)	(0.8, 0.6)	0.046	0.050	0.04	0.024	0.04	0.040	0.05	0.044	0.046	0.054
	(125, 125)	(1, 0.8)	0.046	0.044	0.04	0.040	0.046	0.044	0.048	0.05	0.060	0.056
	(216, 216)	(1, 0.8)	0.066	0.05	0.054	0.050	0.058	0.038	0.050	0.046	0.050	0.052
	(512, 512)	(1, 0.8)	0.04	0.044	0.044	0.034	0.038	0.040	0.054	0.064	0.054	0.068

Table 2.2: Surface comparison: Type 1 error rates.

Func	(n_1, n_2)	(σ_1, σ_2)	TP-Spline	TP-Spline$_+$	TP-Spline$_-$	TE-Spline	TP-Spline.p	TE-Spline.p	T_4	T_1	T_3	T_2
d	(125, 125)	(0.5, 0.3)	0.940	0.944	0.952	0.806	0.944	0.826	0.870	0.784	0.940	0.592
	(216, 216)	(0.5, 0.3)	0.998	0.998	1.000	0.984	0.998	0.990	0.988	0.976	0.994	0.960
	(512, 512)	(0.5, 0.3)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.6, 0.4)	0.792	0.780	0.786	0.656	0.804	0.660	0.742	0.604	0.834	0.400
	(216, 216)	(0.6, 0.4)	0.968	0.958	0.968	0.930	0.972	0.942	0.942	0.902	0.964	0.800
	(512, 512)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
e	(125, 125)	(0.6, 0.4)	0.960	0.962	0.956	0.932	0.968	0.928	0.894	0.918	0.932	0.934
	(216, 216)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998	0.998	0.998
	(512, 512)	(0.6, 0.4)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(0.8, 0.6)	0.730	0.724	0.716	0.618	0.730	0.640	0.650	0.646	0.662	0.706
	(216, 216)	(0.8, 0.6)	0.950	0.946	0.944	0.908	0.944	0.936	0.874	0.880	0.890	0.906
	(512, 512)	(0.8, 0.6)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000
f	(125, 125)	(0.8, 0.6)	0.742	0.742	0.734	0.632	0.736	0.650	0.696	0.684	0.698	0.734
	(216, 216)	(0.8, 0.6)	0.960	0.958	0.962	0.908	0.960	0.926	0.860	0.882	0.894	0.960
	(512, 512)	(0.8, 0.6)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(125, 125)	(1, 0.8)	0.478	0.500	0.490	0.386	0.482	0.388	0.484	0.502	0.500	0.490
	(216, 216)	(1, 0.8)	0.772	0.784	0.784	0.680	0.776	0.736	0.680	0.714	0.722	0.796
	(512, 512)	(1, 0.8)	0.998	0.998	0.998	0.992	0.998	0.996	0.972	0.992	0.990	1.000

Table 2.3: Surface comparison (Cont.): Power.

2.3 Real data application

To illustrate the proposed testing procedure, we analyzed data from an observational study aimed at exploring the relations between pubertal growth and blood pressure development. The fully study protocol was described elsewhere.(Tu et al., 2011, 2014) Briefly, healthy children between 5 and 17 years of age were recruited from schools in Indianapolis, Indiana. Blood pressure, height, and weight were measured from the study participants. The study protocol was approved by a local Internal Review Board. Informed consent was obtained from study participants, or their parents when appropriate.

The current analyses focused on the potentially nonlinear age effects on weight of the study participants. Comparisons were made between height growth curves among four subgroups: white girls, white boys, black girls, and black boys. In order to accommodate the comparison of two groups, a model can be written as

$$Weight_{ij} = g_i(Age_{ij}) + \epsilon_{ij}$$

where i indexes the groups and j the n_i observations within each group. We would like to conduct pairwise comparison among the four groups with $H_0 : g_1 = g_2$ vs $H_1 : g_1 \neq g_2$. The number of observations of the study participants were: 205 black boys, 311 white boys, 232 black girls, and 289 white girls.

We estimated the weight growth curves of the four race-sex groups as part of the preliminary analysis. Scatterplots of weight vs. age are displayed in Figure 2.6(a). P values from the four competing testing methods are presented in Table 2.4. The corre-

sponding estimated regression curves with 95% pointwise confidence intervals using a semiparametric model (Generalized Cross-Validation for selecting smoothing parameter and thin-plate penalized basis function) were presented in Figure 2.6(b). Note that the nonparametric smoothing curves by LOESS gave very close curve estimations compared to our semiparametric estimated ones.

Testing results from the semiparametric spline-based estimating method were consistent with the curve estimations shown in Figure 2.6. The tests showed that the regression function of weight on age were significantly different between white and black girls. Consistently, the black girls gained more weight around ages 12 and 13 than their white peers, but the two curves converged gradually at age 14; and at age 15 the confidence intervals became wider due to reduced sample sizes. These findings were similar with the variable height presented in Table 2.4 and Figure 2.7.

For surface comparisons, we consider the weight as a function of age and height. A model can be written as

$$Weight_{ij} = g_i(Age_{ij}, Height_{ij}) + \epsilon_{ij}$$

where i indexes the groups and j denotes each subjects within each group. We compared simultaneous effects of height and age on weight, among the four race-sex groups. All of the p-values of the four test types are summarized in Table 2.5 and corresponding contour plots are presented in Figure 2.8 and Figure 2.9. No significant differences were detected using the four test types.

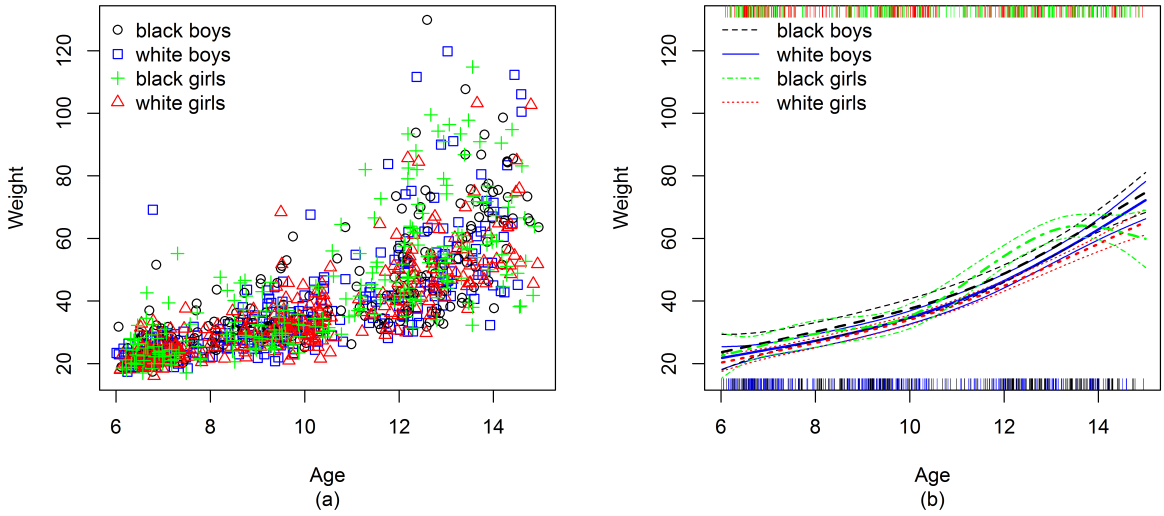

Figure 2.6: (a) Weight over age by groups; (b) Estimated curves of weight on age with pointwise 95% CI by groups

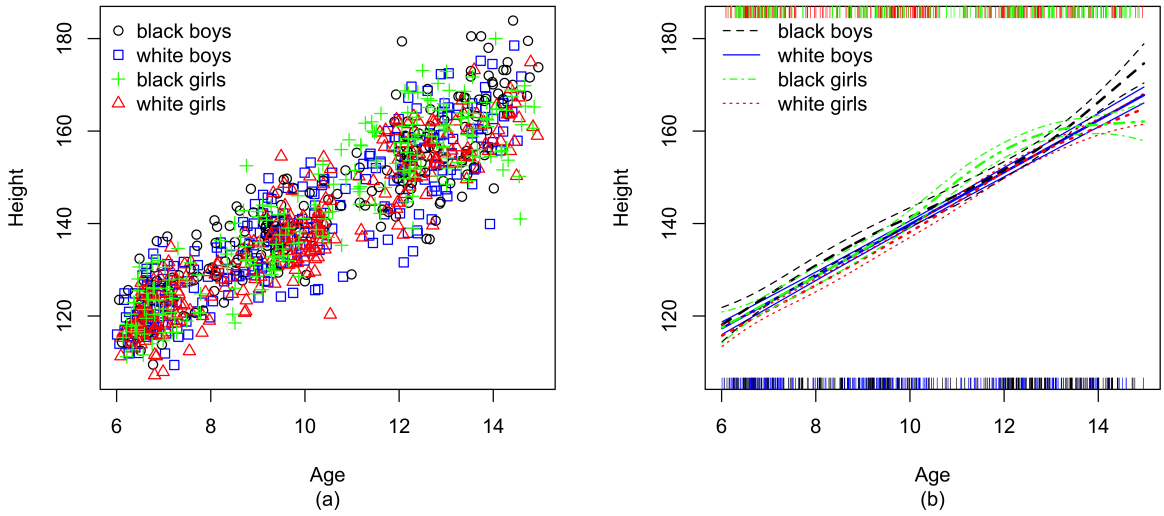

Figure 2.7: (a) Height over age by groups; (b) Estimated curves of height on age with pointwise 95% CI by groups

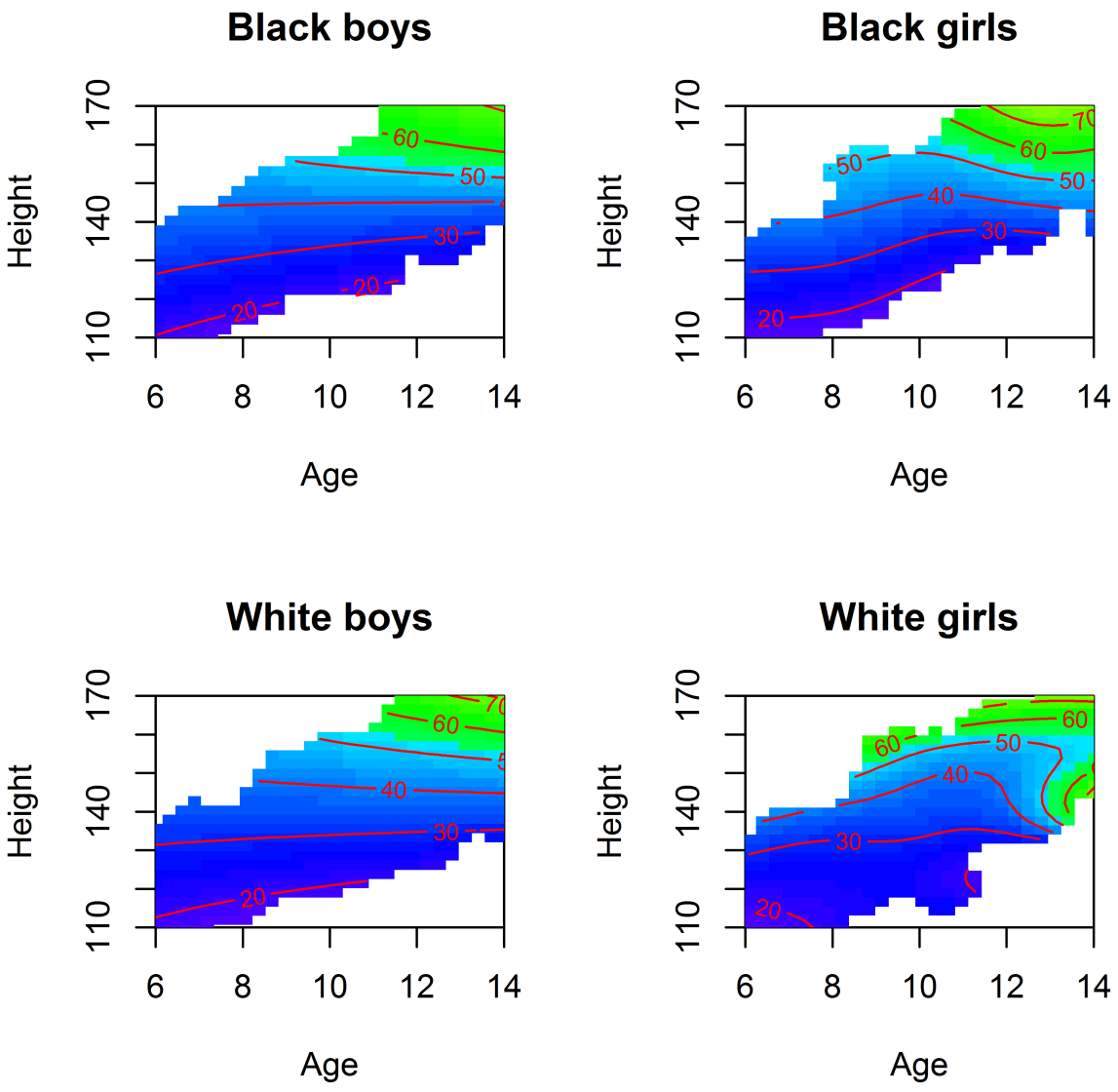

Figure 2.8: Estimated contour plots of weight on height and age by groups

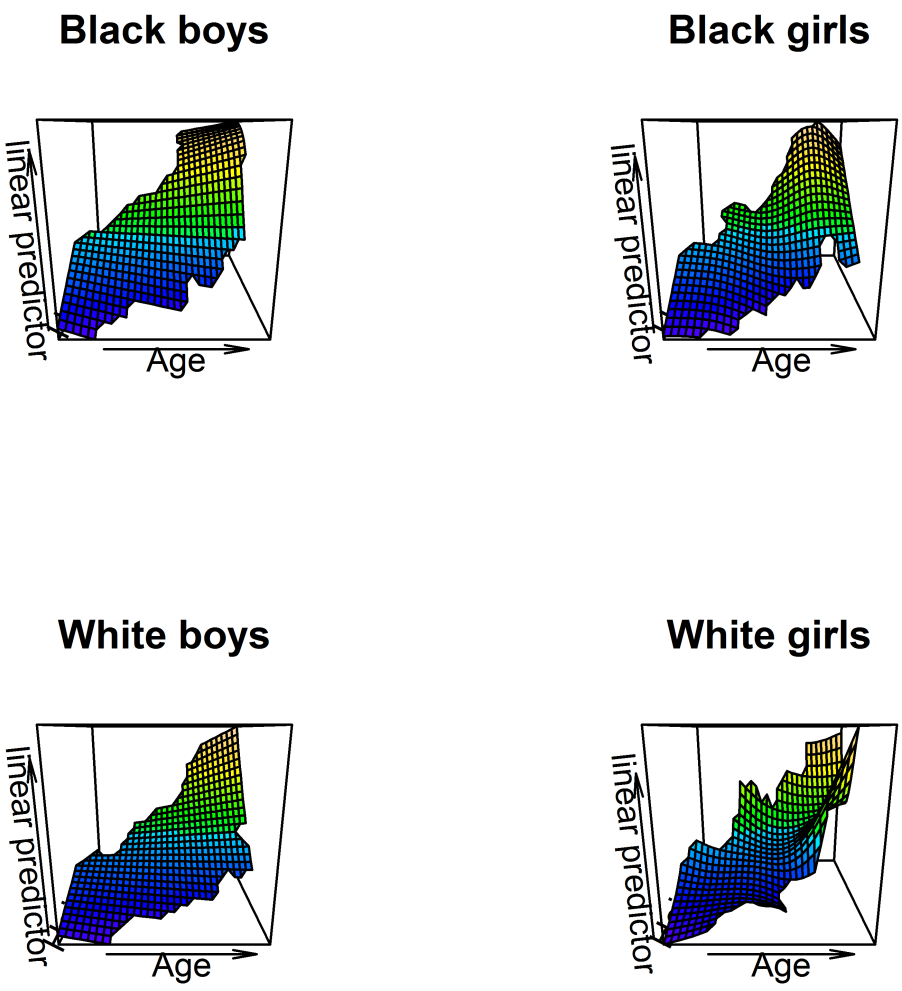

Figure 2.9: Estimated 3D plots of weight on height and age by groups

Endpoint	Group effect	Subset data	$T_{P-spline}$	T_4	T_3	T_2
Weight	Sex	White	0.48	0.18	0.69	0.28
		Black	0.08	0.16	0.55	<0.01
	Race	Girls	<0.01	0.01	0.96	<0.01
		Boys	0.41	0.09	0.69	0.2
Height	Sex	White	0.07	0.04	0.75	0.03
		Black	<0.01	0.01	0.48	<0.01
	Race	Girls	0.01	<0.01	0.85	<0.01
		Boys	0.09	0.01	0.77	0.02

Table 2.4: P-values of testing gender or race differences of nonlinear age association on weight and height

Group effect	Subset data	$T_{P-spline}$	T_4'	T_3'	T_2'
SEX	White	0.49	0.55	0.55	0.77
	Black	0.16	0.65	0.66	0.49
RACE	Male	0.42	0.42	0.46	0.35
	Female	0.34	0.19	0.91	0.06

Table 2.5: P-values of testing gender or race differences of the simultaneous age and height influences on weight

2.4 Discussion

Semiparametric and nonparametric analysis of biomedical data is almost never complete without a proper comparison of the nonlinear relationships between treatment groups or different subpopulations. In studies of low dimensional nonlinear functions, inference typically involves comparisons of curves and surfaces. In parametric analysis where the functions are fully specified, such inference is generally very straightforward and can be carried out in a likelihood-based framework. In nonparametrric or semiparametric analyses, due to the lack of knowledge of the true functional forms of the relationships of interest, analysts are no longer able to rely on likelihood-based tests made

available by standard software packages or functions. In practice, estimation and inference of the functional curves and surfaces are often done separately, due to the lack of integration of estimation and inference tools and common software syntax. We propose a new test for this problem based on an L^2 distance, and show that this test is consistent against any fixed alternative hypothesis. To evaluate the level of statistical significance we provied a set of bootstrap testing methods and an R package to bridge these gaps, which will be illustrated in Chapter 5.

Extensive simulation studies show that, in comparison with the existing methods, the proposed tests have good control of Type 1 error rate and excellent power. Despite our use of computer intensive methods such as wild bootstrapping, the procedures are generally quite efficient. Our own testing of the method with real data suggests the software package is easy to operate and flexible in accommodation of covariates. We will show that the testing procedure possesses the property of consistency, a necessary condition for bootstrap to work in the next chapter. The rate of convergence, however, has not reached the optimal rate of $n^{1/2}$. The empirical evidence from our simulation has nonetheless support the good performance in finite sample situations.

Chapter 3

Asymptotic theory for B-spline-based sieve M-estimation

A fundamental property that ensures the validity of bootstrapping is consistency. Das-Gupta (2008), Beran and Ducharme (1991) further provided the necessary conditions under which a bootstrap estimator achieves consistency.

In the current context, even though the asymptotic normality of the test statistic is difficult to achieve under the null hypothesis, we can still show that the proposed test statistic is consistent. The rate of convergence is still informative of the testing procedure's implementation, especially in the selection of the number of knots. We first rewrite the model in a slightly more general form.

Consider the nonparametric model

$$Y_i = g_0(X_i) + \epsilon_i = g_0(X_i) + \sigma e_i,$$

where g_0 is an unknown smooth function, (Y_i, X_i), $i = 1, \ldots, n$, are i.i.d. random variables independent of the error term $e_i \sim N(0,1)$. For simplicity and without loss of generality assume that the covariate $X_i \in \mathbb{X} = [0,1]$ a.s. (note that any compact subset of $\mathbb{R}$ works).

The estimation of this model can be performed by minimizing the objective function

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i - g(X_i)]^2$$

or, equivalently, by maximizing

$$\mathbb{M}_n(g) \equiv \mathbb{P}_n m_g = 2\mathbb{P}_n(g - g_0)e - \mathbb{P}_n(g - g_0)^2.$$

Direct maximization of the above objective function over the full infinite-dimensional parameter space $\mathcal{G}$ is not possible. Therefore, one can use the sieve M-estimation framework by considering spaces of B-spline functions

$$\mathcal{G}_n(D_n, K_n, m) = \left\{ g_n : g_n(x) = \sum_{k=1}^{K_n+m} \gamma_k B_k^m(x) \in S_n(D_n, K_n, m), x \in [0,1] \right\},$$

where $D_n = \{d_1, \ldots, d_{K_n}\}$ is a set of partition points for the set $[0,1]$, K_n is the number of internal knots with $K_n = O(n^\nu)$, m is the order of the B-spline with $m \geq 1$, $\{\gamma_k\}_{k=1}^{K_n+m}$ is the set of the unknown coefficients or control points for the B-spline, $\{B_k^m(x) : x \in [a,b]\}$ are the basis functions, and $S_n(D_n, K_n, m)$ is the space of polynomial splines on a partition D_n with K_n internal knots and of order m. Then, the sieve estimator $\hat{g}_n$ of g_0 satisfies

$$\mathbb{M}_n(\hat{g}_n) \geq \mathbb{M}_n(g) \text{ for all } g \in \mathcal{G}_n,$$

that is $\hat{g}_n$ maximizes $g \mapsto \mathbb{M}_n(g)$ over the sieve space $\mathcal{G}_n(D_n, K_n, m)$. For simlicity of presentation we will consider the special case of the two-sample comparison. In this work we assume the following regularity conditions:

C1. The errors e have a zero mean and their distribution has subexponential tails. Also, e and the covariate X are independent.

C2. The parameter space $\mathcal{G}_i \ni g_{0,j}$, $i = 1,2$, contains uniformly bounded by $C \geq 1/2$ functions on $[0,1]$, with bounded pth derivatives, for fixed $p \geq 1$, with the first derivative being continuous.

C3. The number of internal knots satisfies $K_n = O(n^{\nu})$, such that

$$\max_{1 \leq k \leq K_n+1} \{d_k - d_{k-1}\} = O(n^{-\nu}).$$

C4. The sample sizes of the two groups satisfy

$$\frac{n_1}{n_1 + n_2} \to \lambda \in (0,1),$$

as $\min(n_1, n_2) \to \infty$.

Theorem 1. *Assume that conditions C1-C4 are satisfied. Then, the proposed test statistic is consistent against any fixed alternative hypothesis.*

Although the proposed test statistic is, by Theorem 1, consistent against any fixed alternative hypothesis , its asymptotic distribution is quite challenging to derive. This stems from the fact that the convergence rate of the B-spline estimator of g_1 and g_2 is

$$d(\hat{g}_{n_i,i}, g_{0,i}) = O_p\left(n^{\frac{p}{1+2p}}\right), \quad i = 1,2,$$

where $d(g_1, g_2) = \{E[g_1(X) - g_2(X)]^2\}^{1/2}$. This convergence rate it is slower than the usual parametric $\sqrt{n}$ rate, even though it is the optimal rate in nonparametric regression. However, the performance of the wild-bootstrap procedure for the calculation of the level of significance can be evaluated though extensive simulation experiments.

Proof of Theorem 1

Consider the L_2-metric

$$d(g_1, g_2) = \|g_1 - g_2\|_{L_2(P)} = \left[P(g_1 - g_2)^2\right]^{\frac{1}{2}},$$

where Pf denotes the expectation of a measurable, data-dependent function f. Now, by condition C1, it is easy to see that the expected objective function is

$$M(g) = Pm_g = -P(g - g_{0,i})^2, \quad i = 1, 2,$$

where $g_{0,i}$ is the true curve for the ith population, which implies the identifiability condition

$$\sup_{g: g \notin G} M(g) < M(g_{0,j})$$

for any neighborhood G_i of $g_{0,i}$. For simplicity of presentation we omit temprorarily the subscript i in the remainder of the proof the substricpt, and note that the same arguments regarding the consistency of the B-spline estimator hold for both groups. Next, similar to Kim et al. (2017), define a linear operator map $\mathscr{Q}$ from $\mathscr{G}$ to the sieve space $\mathscr{G}_n$, as:

$$\mathscr{Q}(\psi) = \sum_{k=1}^{K_n+m} \phi_k(\psi) B^m(x),$$

for any $\psi \in \mathscr{G}$, where $\{\phi_k\}_{k=1}^{K_n+m}$ are linear functionals in $L_\infty(\mathbb{X})$. Now, define $g_n(x) = \mathscr{Q}(g_0)$. Similar arguments to those used in Kim et al. (2017) lead to the inequality

$$\|g_n - g_0\|_{L_\infty(\mathbb{X})} \leq O(K_n^{-p}),$$

which also implies that

$$\|g_n - g_0\|_{L_2(P)} \leq O(K_n^{-p}). \tag{3.1}$$

Now, it is straightforward to see that

$$d(\hat{g}_n, g_0) \leq d(\hat{g}_n, g_n) + d(g_n, g_0). \tag{3.2}$$

To show the convergence of the first term in the right side of (3.2) to 0, we further define $\mathbb{M}_n(g)$ is the empirical objective function based on the data and $\mathbb{P}_n f(X) = n^{-1}\sum_{i=1}^{n} f(X_i)$, following which

$$\begin{aligned} \sup_{g\in\mathscr{G}_n} |\mathbb{M}_n(g) - M(g)| &\equiv \|\mathbb{M}_n(g) - M(g)\|_{\mathscr{G}_n} \\ &\lesssim \|\mathbb{P}_n(g-g_0)e\|_{\mathscr{G}_n} + \|(\mathbb{P}_n - P)(g-g_0)^2\|_{\mathscr{G}_n}. \end{aligned}$$

By the conditions C1, C2, and the law of large numbers it follows that $\|\mathbb{P}_n(g-g_0)e\|_{\mathscr{G}_n} = o_p(1)$. For the second term consider the class of functions $\mathscr{F}_n = \{(g-g_0)^2 : g \in \mathscr{G}_n\}$. A calculation by Shen and Wang (1994) implies that $N_{[]}(\epsilon, \mathscr{G}_n, L_1(P)) \leq (1/\epsilon)^{c(K_n+m)}$. Based on the set of ϵ-brackets $\{[l_j, u_j] : j = 1, \ldots, (1/\epsilon)^{c(K_n+m)}\}$ in $L_1(P)$ for $\mathscr{G}_n$, we can construct a set of ϵ'-brackets $\{[(l_j - g_0)^2, (u_j - g_0)^2] : j = 1, \ldots, (1/\epsilon)^{c(K_n+m)}\}$ in $L_1(P)$ for $\mathscr{F}_n$. Therefore, $N_{[]}(\epsilon, \mathscr{F}_n, L_1(P)) < \infty$ for any $\epsilon > 0$ which implies that $\|(\mathbb{P}-P)(g-g_0)^2\|_{\mathscr{G}_n} \overset{as*}{\to} 0$, and thus $\|\mathbb{M}_n(g) - M(g)\|_{\mathscr{G}_n} = o_p(1)$. This fact along with the inequality

$$M(g) - M(g_n) \leq -\frac{1}{4}P(g-g_n)^2,$$

for any g with $P(g-g_n)^2 \geq 4P(g_n-g_0)^2$ (van der Vaart and Wellner, 1996) implies that $d(\hat{g}_n, g_n) = o_p(1)$. Now, the second term on the right side of (3.2) is $o(1)$ by condition C3 and inequality (3.1) and this leads to $d(g_n, g_0) = o(1)$ and, therefore, $d(\hat{g}_n, g_0) \xrightarrow{p} 0$.

Now, for the rate of convergence consider the key inequality

$$E^\star \sup_{P(g-g_n)^2 \leq \delta^2, g \in \mathcal{G}_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g-g_n)(X_i) e_i \right| \lesssim \tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P)) \left[1 + \frac{\tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P))}{\delta^2 \sqrt{n}} \right]$$

holds with $\mathcal{G}_n(\delta) = \{g : g \in \mathcal{G}_n, d(g, g_n) < \delta\}$, given in p. 335 in van der Vaart and Wellner (1996). The calculation by Shen and Wang (1994) implies that the ϵ-bracketing number for the class $\mathcal{G}_n(\delta)$ is bounded by $(\delta/\epsilon)^{c(K_n+m)}$. Therefore,

$$\tilde{J}_{[]}(\delta, \mathcal{G}_n(\delta), L_2(P)) = \int_0^\delta \sqrt{1 + c(K_n+m) \log\left(\frac{\delta}{\epsilon}\right)} d\epsilon \leq c(K_n+m)^{1/2} \delta.$$

Thus, the key function $\phi_n(\delta)$ given in Theorem 3.4.1. in van der Vaart and Wellner (1996) is

$$\phi_n(\delta) = (K_n+m)^{1/2} \delta + \frac{(K_n+m)}{\sqrt{n}}.$$

Now, after some algebra we conclude that

$$n^{2pv} \phi_n\left(\frac{1}{n^{pv}}\right) \leq \sqrt{n}$$

if $pv \leq (1-v)/2$. Thus, if the rate $r_n = \min(pv, (1-v)/2)$ then

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) \leq \sqrt{n},$$

and $\mathbb{M}_n(\hat{g}_n) - \mathbb{M}_n(g_0) \geq -O_p(r_n^{-2})$. Now if $\nu = 1/(1+2p)$, Theorem 3.4.1 in van der Vaart and Wellener (1996) and (3.1) and (3.2) imply that

$$d(\hat{g}_n, g_0) = O_p(n^{-\frac{p}{1+2p}}).$$

Next, consider the test statistic based on two independent samples of size n_1 and n_2

$$\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}[\hat{g}_1(\mathbf{x}_{ij}) - \hat{g}_2(\mathbf{x}_{ij})]^2 \equiv \mathbb{P}_N(\hat{g}_1 - \hat{g}_2)^2,$$

where $N = n_1 + n_2$.

Let us consider the consistency of the test statistic to the true L^2 distance between the two curves under the probability measure P underlying X_{ij}, $i = 1,2$, $j = 1,\ldots,n_i$. A straightforward expansion leads to

$$\begin{aligned}\left|\mathbb{P}_N(\hat{g}_1 - \hat{g}_2)^2 - P(g_1 - g_2)^2\right| \quad &\leq \quad \left|\mathbb{P}_N[(\hat{g}_1 - \hat{g}_2)^2 - (g_1 - g_2)^2]\right| \\ &\quad + \left|(\mathbb{P}_N - P)(g_1 - g_2)^2)\right|. \qquad (3.3)\end{aligned}$$

The second term in (3.3) is $o_p(1)$, as a consequence of the condition C2 and the law of large numbers. Now see that

$$\begin{aligned}
\mathbb{P}_N(\hat{g}_i - g_i)^2 &= \mathbb{P}_N \left\{ \sum_{k=1}^{K_{n_i}+m} [\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)] B_k^m + \left[\sum_{k=1}^{K_n+m} \phi_{k,i}(g_i) B_k^m - g_i \right] \right\}^2 \\
&\equiv \mathbb{P}_N \left\{ \sum_{k=1}^{K_{n_i}+m} [\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)] B_k^m + (g_{n_i,i} - g_i) \right\}^2 \\
&\leq 2 \left\{ \sum_{k=1}^{K_{n_i}+m} [\hat{\phi}_{k,i}(g_i) - \phi_{k,i}(g_i)] \right\}^2 \mathbb{P}_N \left(\max_{k=1,\ldots,K_{n_i}+m} B_k^m \right)^2 \\
&\quad + 2P(g_{n_i,i} - g_i)^2 + o_p(1).
\end{aligned}$$

By the uniform boundedness of the B-spline bases functions and the consistency of the estimator of the control points $\hat{\phi}_{k,i}(g_i)$ from the fact that $d(\hat{g}_{i,n}, g_{i,n}) = o_p(1)$ shown above, it follows that the first term in the right side of the above inequality is $o_p(1)$. Also, the second term in the right side of the above inequality is $o_p(1)$, as it was argued above. Therefore,

$$\mathbb{P}_N(\hat{g}_i - g_i)^2 = o_p(1), \quad i = 1, 2. \tag{3.4}$$

(3.4) can be written into

$$\begin{aligned}
\left| \mathbb{P}_N[(\hat{g}_1 - \hat{g}_2)^2 - (g_1 - g_2)^2] \right| &\leq \sum_{i=1}^{2} \left| \mathbb{P}_N(\hat{g}_i^2 - g_i^2) \right| + \left| \mathbb{P}_N \hat{g}_1 (\hat{g}_2 - g_2) \right| \\
&\quad + \left| \mathbb{P}_N g_2 (\hat{g}_1 - g_1) \right| \\
&\equiv \sum_{i=1}^{2} A_{N,i} + B_N + C_N
\end{aligned} \tag{3.5}$$

It is not hard to see that the first term of (3.5) becomes

$$\begin{aligned}
\sum_{i=1}^{2} A_{N,i} &\leq \sum_{i=1}^{2} \mathbb{P}_N \left(\hat{g}_i - g_i\right)^2 + \sum_{i=1}^{2} \mathbb{P}_N \left|2g_i(\hat{g}_i - g_i)\right| \\
&\leq \sum_{i=1}^{2} \mathbb{P}_N (\hat{g}_i - g_i)^2 + K \sum_{i=1}^{2} \mathbb{P}_N \left|2g_i(\hat{g}_i - g_i)\right| \\
&\leq \sum_{i=1}^{2} \mathbb{P}_N (\hat{g}_i - g_i)^2 + K \sum_{i=1}^{2} [\mathbb{P}_N (\hat{g}_i - g_i)^2]^{1/2} \\
&= o_p(1),
\end{aligned}$$

where K represents a generic constant with $K \in (0, \infty)$. By C2 and (3.5), $i = 1, 2$, we have that for a suffiently large N

$$(B_N + C_N) \leq K \sum_{i=1}^{2} \mathbb{P}_N (\hat{g}_i - g_i)^2 = o_p(1).$$

Therefore,

$$\mathbb{P}_N (\hat{g}_1 - \hat{g}_2)^2 \xrightarrow{p} P(g_1 - g_2)^2.$$

This result along with Lemma 14.15 of Van der Vaart (2000) leads to the consistency of the proposed test against every fixed alternative hypothesis with $g_1 \neq g_2$. Thus, the proof of Theorem 1 is complete.

Chapter 4

Curve and Surface Comparison Using Longitudinal Data

4.1 A test based on Semiparametric Mixed Models

4.1.1 Semiparametric mixed effect model

In analysis of longitudinal data, subject-specific correlations have to be accounted for. Semiparametric mixed models allow for more flexible specification of the within-subject variance-covariance structure (Zhang et al., 1998). Let Y_{ijl} be the ith group subject j at lth visit, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n_i$, and $l = 1,2,\ldots,q_{ij}$, where k is the number of groups, n_i is the number of subjects in each group, and q_{ij} are the follow-up visits for the jth individual in group i. Here Y_{ijl} is a continuous outcome. A multivariate semiparametric mixed model can be expressed as

$$Y_{ijl} = g_i(x_{1ijl}, x_{2ijl}) + \mathbf{Z}_{ijl}^T \mathbf{b}_{ij} + \epsilon_{ijl} \tag{4.1}$$

where (x_{1ijl}, x_{2ijl}) are the main explanatory variables of interest, $g_i(x_{1ijl}, x_{2ijl})$ is the nonparametric function of group i, and $\mathbf{b}_{ij} = (\mathbf{b}_{i1},\ldots,\mathbf{b}_{ip})$ is the subject-specific random effect vector following a multivariate normal distribution with mean zero, $N_p(0, D_i(\phi))$. Here $D_i(\phi)$ is the unknown positive definite covariance matrix with parameter ϕ, and $\epsilon_i \sim N(0, \sigma_i^2 \mathbf{I})$, where $\mathbf{I}$ is the identity matrix of dimension $N = \sum_{i=1}^{k} n_i$.

The semiparametric mixed model (4.1) could be expressed as a regular linear mixed model, by treating the linear effects $\mathbf{X_i}\beta$ as fixed and spline coefficients as random

$\mathbf{Z_i b}$. First, let us consider a univariate smoothing component, followed by a bivariate smoothing component in the form of tensor product splines.

We first consider a univariate spline, e.g. $g(\mathbf{x}) = \sum_{j=1}^{J} b_j(\mathbf{x}\beta_j)$ with associated a smoothing parameter, $J(g) = \beta^{\mathbf{T}}\mathbf{S}\beta$, where $\mathbf{S}$ is a positive semi-definite matrix of coefficients (semi-definite because most penalties treat some space of functions as having zero wiggliness). Given $(y_i, \mathbf{x}_i)$, a model matrix $\mathbf{X}^g$ could be produced so that $\mathbf{X}^g\beta$ is a vector of $g(\mathbf{x}_i)$ values. To estimate g, we first use eigen- decomposition, $\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U}$ is an orthogonal matrix, the columns of which are the eigenvectors of $\mathbf{S}$, and $\mathbf{D}$ is a diagonal matrix containing corresponding eigenvalues. Let $\mathbf{D}_+$ be the smallest sub-matrix of $\mathbf{D}$ containing all the strictly positive eigenvalues. Then by reparameterization, the new coefficient vector could be written $(\mathbf{b}_R^T, \beta_F^T) = \mathbf{U}^T\beta$, where β_F is unpenalized and is of dimension M. It is clear that $\beta^{\mathbf{T}}\mathbf{S}\beta = \mathbf{b}_R^T\mathbf{D}_+\mathbf{b}_R$. By partitioning the eigenvector matrix $\mathbf{U}$ into $[\mathbf{U}_R : \mathbf{U}_F]$, where $\mathbf{U}_F$ has M columns, we define $\mathbf{X}_F = \mathbf{X}^g\mathbf{U}_F$ and $\mathbf{X}_R = \mathbf{X}^g\mathbf{U}_R$. It then follows that the mixed model representation of the smoothing function can be written as $\mathbf{X}_F\beta_F + \mathbf{X}_R\beta_R$, where $\mathbf{b}_R \sim N(\mathbf{0}, \mathbf{D}_+^{-1}/\lambda)$. Here λ and β are the fixed parameters to be estimated. The model is now expressed as a standard generalized linear mixed model (GLMM) by appending the columns of $\mathbf{X}_F$ to the fixed effect design matrix, appending the columns of $\mathbf{Z}$ to the random effects design matrix, and specifying the given random effect covariate matrix.

When we had multiple smoothing parameters and present tensor product smooths, the only change was that the covariance matrix for β requires re-parameterization similar to the part described in the tensor product section in Chapter 2 Section 2.1.

4.1.2 Bootstrap technique for correlated data to obtain p-value

The same testing procedure can be modified and extended to situations of correlated data. A key feature of the modification is the preservation of the correlation structure existing within each subject. The following algorithm is a natural extension and it relies on a Cholesky decomposition of the estimated covariance matrix (McMurry and Politis, 2010).

We present the following algorithm.

1. Estimate $\hat{g}_i(\mathbf{x})$ of the two groups separately using semiparametric mixed effect models, and compute the test statistic $T_{splcorr}$.
2. Estimate the common regression function $\hat{g}(\mathbf{x})$ by using the combined sample.
3. For $i = 1,\ldots,k$, estimate the corresponding covariance matrix $\hat{\mathbf{R}}_i$ for each group based on the fitted individual semiparametric model and further ascertain the residuals $\hat{\eta}_i = (\hat{\eta}_{i11}, \hat{\eta}_{i12}, \ldots, \hat{\eta}_{inn_i})$.
4. Perform a Cholesky decomposition $\hat{\mathbf{L}}_i$ such that $\hat{\mathbf{R}}_i = \hat{\mathbf{L}}_i \hat{\mathbf{L}}_i^T$ and obtain

$$\hat{\mathbf{e}}_i = (\hat{e}_{i11}, \hat{e}_{i12}, \ldots, \hat{e}_{inn_i}) = \hat{\mathbf{L}}_i^{-1} \hat{\eta}_i;$$

then calculate the "whitened" residual $\tilde{\mathbf{e}}_i$ by computing $\tilde{\mathbf{e}}_i = \hat{\mathbf{e}}_i - \frac{1}{n}\sum_{j=1}^{n_i} \hat{\mathbf{e}}_i$.
5. Draw the "whitened" bootstrap residuals $\tilde{\mathbf{e}}_i$ and generate a bootstrap sample by computing $\tilde{\eta}_i = \hat{\mathbf{L}}_i \tilde{\mathbf{e}}_i$, and then compute the bootstrap observations by setting $Y^*_{ijl} = \hat{g}(\mathbf{x_{ijl}}) + \tilde{\eta}^*_{ijl}$.
6. Calculate the test statistic T^*_{spline} using the general bootstrap sample. Repeat steps 3-5 B times to approximate the distribution of the statistic with B bootstrap

replicates and then determine the quantiles of the test statistic under the null hypothesis.

4.2 Simulation studies for curve comparison with repeated measurements

We considered the following models for correlated data

$$Y_{ijl} = g_{id}(X_{ijl}) + b_{ij} + \epsilon_{ijl},$$

where $i = 1,2; j = 1,\ldots,n_i; l = 1,2,3$. As previously presented in Equation (2.3), we used Y_{ijl} to indicate the measure on the lth occasion from the jth subject in the ith group. Values of independent variable x_{ijl} were generated from independent Unif[0,1]. Values of the dependent variables Y_{ij} were generated from the above equations with random effect $b_i \sim N(0,\sigma_i')$ and the i.i.d random error $\epsilon_{ijl} \sim N(0,\sigma_i)$. We used the same regression functions defined in Section 2.2.1, where the two curves gradually grew apart with increasing d (as plotted in Figure 2.2).

The simulation was performed under the following parameter settings: (1) d=0, 1, 2; (2) three sample size settings (n_1,n_2) = (50, 60), (100, 120), and (150, 160) and all with three repeated measures; (3) three different combinations of standard errors for the random intercept and the i.i.d random variable as $(\sigma_1',\sigma_2',\sigma_1,\sigma_2)$= (0.2, 0.15, 0.04, 0.05), (0.2, 0.15, 0.10, 0.12), and (0.25, 0.20, 0.10, 0.12).

We used penalized semiparametric mixed regression to estimate the curves. The rejection probabilities under the null based on 1000 simulation runs are shown in Table 4.1. The significance levels of all the tests were set to 0.05. At $d = 0$, type I error rates were in general well controlled across different sample sizes and variance settings. As d

increased, the power for rejecting the null grew. The rejection probability increased with smaller variances and larger sample sizes.

For comparison with Zhang et al.'s scaled chi-square test, we performed comparisons using 1) the same $\mathbf{x}$ for two groups; 2) slightly different $\mathbf{x}$ ($\mathbf{x}_2 = \mathbf{x}_1$ + Unif(0,0.05)); and 3) completely random and independent $\mathbf{x}_1, \mathbf{x}_2$ for two groups. Simulation was repeated 200 times under each scenario and the results are shown in Table 4.2.

When the two groups shared the same $\mathbf{x}$, the scaled χ^2 test had Type I error rates that were much smaller than the nominal level. The power was generally much lower as well. As we introduced a slight difference in $\mathbf{x}$ between the two groups – less than 5% of the range of $\mathbf{x}$, the scaled χ^2 test had a dramatically lower power in contrast to a more stable performance of the proposed method. Moreover, when $\mathbf{x}_1$ and $\mathbf{x}_2$ were simulated independently and randomly, the power of the scaled χ^2 method was completely diminished.

d	(n_1, n_2)	σ_1'	σ_2'	σ_1	σ_2	$T_{splinecorr}$
0	(50,60)	0.20	0.15	0.04	0.05	0.057
	(50,60)	0.20	0.15	0.10	0.12	0.058
	(50,60)	0.25	0.20	0.10	0.12	0.062
	(100,120)	0.20	0.15	0.04	0.05	0.066
	(100,120)	0.20	0.15	0.10	0.12	0.072
	(100,120)	0.25	0.20	0.10	0.12	0.068
	(150,160)	0.20	0.15	0.04	0.05	0.048
	(150,160)	0.20	0.15	0.10	0.12	0.041
	(150,160)	0.25	0.20	0.10	0.12	0.053
1	(50,60)	0.20	0.15	0.04	0.05	0.426
	(50,60)	0.20	0.15	0.10	0.12	0.383
	(50,60)	0.25	0.20	0.10	0.12	0.212
	(100,120)	0.20	0.15	0.04	0.05	0.968
	(100,120)	0.20	0.15	0.10	0.12	0.784
	(100,120)	0.25	0.20	0.10	0.12	0.516
	(150,160)	0.20	0.15	0.04	0.05	1.000
	(150,160)	0.20	0.15	0.10	0.12	0.941
	(150,160)	0.25	0.20	0.10	0.12	0.804
2	(50,60)	0.20	0.15	0.04	0.05	1.000
	(50,60)	0.20	0.15	0.10	0.12	0.994
	(50,60)	0.25	0.20	0.10	0.12	0.956
	(100,120)	0.20	0.15	0.04	0.05	1.000
	(100,120)	0.20	0.15	0.10	0.12	1.000
	(100,120)	0.25	0.20	0.10	0.12	1.000
	(150,160)	0.20	0.15	0.04	0.05	1.000
	(150,160)	0.20	0.15	0.10	0.12	1.000
	(150,160)	0.25	0.20	0.10	0.12	1.000

Table 4.1: Type 1 error rates and power of comparisons with correlated data.

			$\mathbf{x_2} = \mathbf{x_1}$		$\mathbf{x_2} = \mathbf{x_1} + U(0, 0.05)$		Random $\mathbf{x_1}, \mathbf{x_2} \sim U(0,1)$	
d	(n_1, n_2)	$(\sigma_1', \sigma_2', \sigma_1, \sigma_2)$	Scaled χ^2	$T_{splinecorr}$	Scaled χ^2	$T_{splinecorr}$	Scaled χ^2	$T_{splinecorr}$
0	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.000	0.045	0.010	0.075	0	0.070
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.005	0.070	0.005	0.070	0	0.065
	(100,100)	(0.20, 0.15, 0.04, 0.05)	0.005	0.035	0.025	0.070	0	0.065
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.005	0.050	0.010	0.035	0	0.030
1	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.035	0.360	0.005	0.330	0	0.405
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.005	0.180	0.000	0.165	0	0.185
	(100,100)	(0.20, 0.15, 0.04, 0.05)	0.250	0.955	0.000	0.900	0	0.965
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.070	0.535	0.000	0.400	0	0.430
2	(50,50)	(0.20, 0.15, 0.04, 0.05)	0.765	1.000	0.020	1.000	0	1.000
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.270	0.890	0.010	0.885	0	0.935
	(100,100)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.025	1.000	0	1.000
	(100,100)	(0.25, 0.20, 0.10, 0.12)	0.840	1.000	0.005	1.000	0	1.000
3	(50,50)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.460	1.000	0	1.000
	(50,50)	(0.25, 0.20, 0.10, 0.12)	0.915	1.000	0.185	1.000	0	1.000
	(100,100)	(0.20, 0.15, 0.04, 0.05)	1.000	1.000	0.905	1.000	0	1.000
	(100,100)	(0.25, 0.20, 0.10, 0.12)	1.000	1.000	0.565	1.000	0	1.000

Table 4.2: Compare the rejection probability between Zhang DW et al's scaled χ^2 testing method with our method

4.3 Data Analysis

The same data source was used to illustrate the testing procedure with correlated data. All the visits were kept for analysis in this section. Our goal was to compare the growth curve over age between boys and girls. The p-value of comparing the age association with height was significant (p<0.001). Scatterplots of height vs. age are displayed in Figure 4.1. The corresponding estimated regression curves with 95% pointwise confidence intervals using a semiparametric mixed model are presented in Figure 4.1.

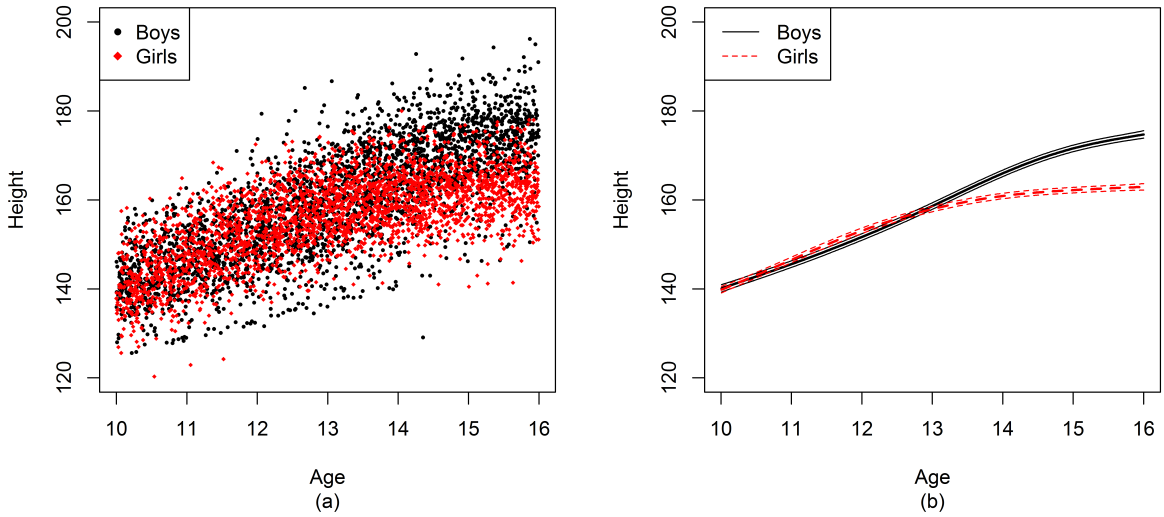

Figure 4.1: (a) Height over age by groups; (b) Estimated fixed effect regression curves of height on age with pointwise 95% CI by groups

4.4 Summary

The simulation results indicated that the proposed testing method with the cluster bootstrap algorithm had a satisfactory performance for longitudinal data. Zhang et al.'s ex-

isting scaled χ^2 methods requires the same measurements of the predictor, and are too restrictive to be used in practical data analysis.

Furthermore, as repeated-measurement datasets are commonly seen in practice where the semiparametric mixed modeling techniques have been well-adopted, our testing method based on these existing techniques are familiar to most data analysts, and thus might be more easily acceptable.

Chapter 5

Software Development

5.1 Package “gamm4.test" in R

We submitted an R package titled `gamm4.test` to CRAN to make the proposed testing procedures available to analytical practitioners.The two main testing functions are `gam.grptest` for comparisons of nonlinear functions with cross-sectional data, and `gamm4.grptest` for comparisons involving correlated data. The main features of this packages include:

1. Syntax that is consistent with `mgcv` and `gamm4`, and packages that are often used for fitting semiparametric regression models. Analysts familiar with those packages can specify the model structures in a familiar syntax and perform desired comparisons using the two testing functions.
2. Use of parallel computing with functions `gam.grptest` and `gamm4.grptest` for an enhanced the computational efficiency.
3. Plotting features within the package for more convenient visual examination and comparison of the fitted curves and surfaces. Graphics are produced by using the R package `plotly`. Setting argument `type` equals to `"plotly.persp"` returns a 3-D plot that users can interact with. Finally, option `test.statistic = TRUE` produces the empirical distribution of the test statistic under the null hypothesis of equal regression functions.

5.1.1 Installation

One can install the latest version of this package from Github as below:

```
if (!require('devtools')) install.packages('devtools')
devtools::install_github('zhaoshi169/gamm4.test',force=TRUE)
```

5.1.2 Functions and Examples

Cross-sectional data: An example

Our first example compares the weight-for-age curves for boys and girls from the pubertal growth study. Because we only use the baseline measurement, the data are cross-sectional.

```
R> library("gamm4.test")
R> data("outchild")
R> child <- outchild[order(outchild$SID,outchild$age),]
R> bs <- aggregate(.~SID, child, FUN=head, 1)

R> childcur <- bs[,c("SEX","WEIGHT","age")]
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="cr"),
      test=~SEX,data=childcur)
R> test.grpsex1
```

The output from the program thus far is as follows:

```
        Test the equality of curves based on L2 distance

Comparing 2 semiparametric regression curves
Penalized semiparametric regression is used for curve fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.
```

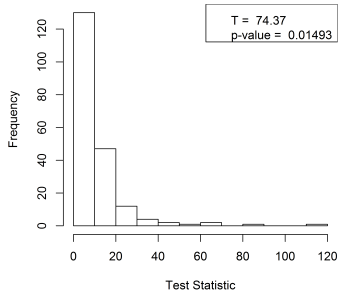

Figure 5.1: Empirical distribution of the test statistic under the null hypothesis

```
Null hypothesis: there is no difference between the 2 curves.
T =  71.92      p-value =  0.01493
```

Printing the object returned from `gam.grptest()` provides the test statistic, T, the p-value as calculated using the bootstrap procedure. A visual depiction of the test in the form of an histogram depicts the distribution of the bootstrap sample returned by `gam.grptest()`. The argument, however, `test.stat.type` may be changed to return a density curve for the bootstrap estimates, instead of an histogram.

```
plot(test.grpsex1, test.statistic=TRUE).
```

In Figure 5.1 we request a histogram by using

```
plot(test.grpsex1, test.statistic=TRUE, test.stat.type="hist")
```

Function `gam.grptest` is the main function for comparing the curves. Arguments `bs` and `k` may be used in the model formula in order to change basis functions and the number of knots used in construction of the spline function. If no value for `k`

is provided, a penalized semiparametric model estimation with the default number of knots will be used. Below are some sample commands:

```
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="tp"),test=~SEX,
      data=childcur)
      #penalized thin-plate spline basis
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="tp", k=5), test=~SEX,
       data=childcur)
      #thin-plate spline basis with five equally spaced knots
      #over the range of variable age
```

Argument `N.boot` specifies the number of bootstrap samples. The default value is `N.boot = 200`. `parallel=TRUE` requests parallel computing and distributes the computational burden to all available CPU cores. Several dependent packages are involved in the parallization process. First, the R Package `parallel` is used to determine the number of cores and initiate cluster (which is a collection of “workers” that will be doing the job). Then we "register" the cluster using package `doParallel`. A function is then written prior to model fitting and test statistic calculation with each bootstrap sample. Finally, with the `foreach` package, which is popular for creating "for loop and lapply", the function is implemented on bootstrap samples and creates a collection of T^*.

```
R> test.grpsex1 <- gam.grptest(WEIGHT~s(age, bs="cr"),test=~SEX,
      data=childcur, N.boot=300, parallel= TRUE)
R> test.grpsex1
```

The following code produces a plot of the estimated curves with a 95% pointwise confidence interval. The plot is shown in Figure 5.2.

```
R> plot(test.grpsex1)
```

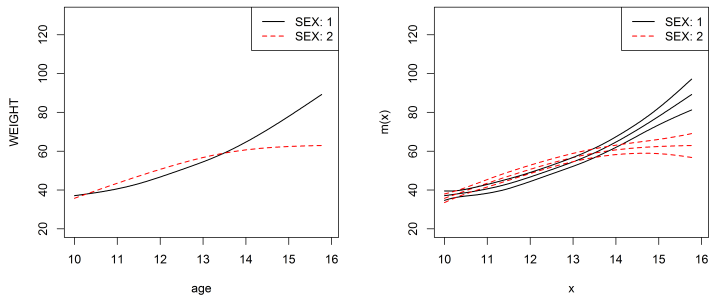

Figure 5.2: Weight over age by sex and pointwise 95% CI by sex

```
R> plot(test.grpsex1, se.est=TRUE)
```

Similarly, one could use the function `gam.grptest` for comparison of surface functions. In the pubertal growth example, we express the body weight as a function of age and height, i.e., $WEIGHT = f(HEIGHT, age)$. The following code produces a comparison of the function f between boys and girls.

```
R> childsurf <- bs[,c("SEX","HEIGHT","WEIGHT","age")]
R> test.grpsex2 <- gam.grptest(WEIGHT~s(HEIGHT,age),test=~SEX,
      data=childsurf)
R> test.grpsex2
R> plot(test.grpsex2)
R> plot(test.grpsex2,type="persp",theta=-35,phi=40)
R> plot(test.grpsex2,type="plotly.persp")
```

with the following output:

```
        Test the equality of surfaces based on L2 distance


Comparing 2 semiparametric regression surfaces
Penalized semiparametric regression is used for surface fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.
```

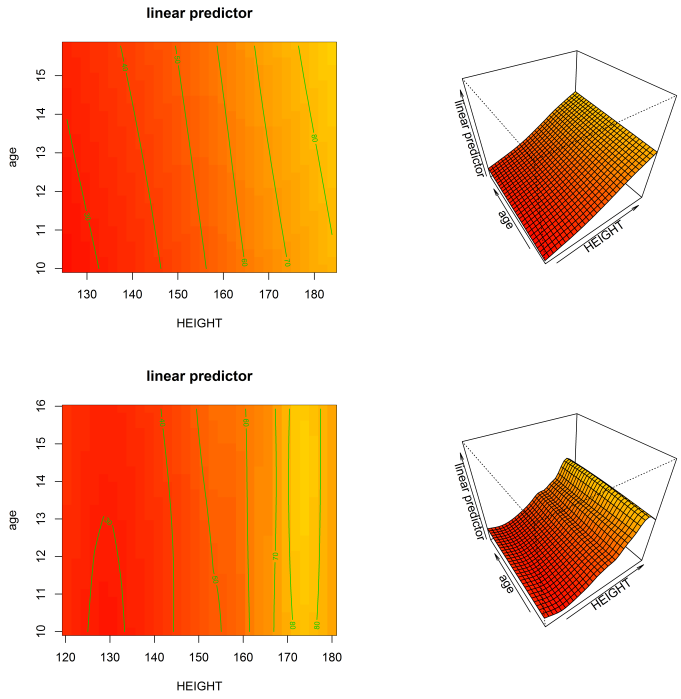

Figure 5.3: Estimated contour and 3D plots of weight on height and age by gender

```
Null hypothesis: there is no difference between the 2 surfaces.
T =  20.92      p-value =  0.4179
```

Setting argument `type=plotly.persp` will generate an interactive 3-D plot using the R package `plotly`. The generated plots are shown in Figure 5.3 and plotly output are uploaded on `https://zhaoshi169.github.io/chap5plotlysuf.html`.

We created a new function called `T.L2c`, which builds off the existing `T.L2` function from package `fANCOVA`. Our new function allows for two group comparisons.

```
R> n1 <- 200
R> x1 <- runif(n1,min=0, max=3)
R> sd1 <- 0.2
R> e1 <- rnorm(n1,sd=sd1)
```

```
R> y1 <- sin(2*x1) + cos(2*x1) + e1
R>
R> n2 <- 120
R> x2 <- runif(n2, min=0, max=3)
R> sd2 <- 0.25
R> e2 <- rnorm(n2, sd=sd2)
R> y2 <- sin(2*x2) + cos(2*x2) + x2 + e2
R>
R> dat <- data.frame(rbind(cbind(x1,y1,1), cbind(x2,y2,2)))
R> colnames(dat)=c('x','y','group')
R>
R> T.L2c(formula=y~x,test=~group,data=dat)
R> gam.grptest(y~s(x,bs="cr"), test=~group, data=dat, parallel=TRUE)
R> library(fANCOVA)
R> T.aov(dat$x, dat$y, dat$group)
R> T.var(dat$x, dat$y, dat$group)
```

In this simple example, all testing methods correctly reject the null hypothesis. But as we have shown in the simulation studies, different testing methods do have different operating characteristics.

Analysis of correlated data

In this example, repeated measures are included in the analysis with the goal of comparing the average growth rates over age between boys and girls.

The R code for data preparation and testing is presented as below.

```
R> data("outchild")
R> child.rep <- outchild[(outchild$age<16 &outchild$age>10),]
```

```
R> child.reptest1 <-  gamm4.grptest(HEIGHT~s(age,bs="cr"),
      random=~(1|SID),test=~SEX,data=child.rep)
R> child.reptest1
R> plot(child.reptest1,test.statistic=TRUE)
R> plot(child.reptest1)
```

The output is as follows:

```
  Test the equality of curves based on L2 distance


Comparing 2 semiparametric regression curves
Penalized semiparametric regression mixed model is used for curve fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.


Null hypothesis: there is no difference between the 2 curves.
T =  33.91     p-value =  0.004975
```

Note that the familiar looking syntax employed by the `gamm4.test` package is consistent with that of `mgcv` and `gamm4`. Specifically, the R formula interface is used throughout, and random effects are specified with the random argument, similar to the existing R packages.

As expected, the height-for-age growth curves are significantly different between the sexes ($p<0.01$). The empirical distribution of the test statistic under the null hypothesis shows that the value of the test statistic, $T = 33.91$, is located to the far right of the plotted range. At the same time, the 95% pointwise confidence bands were relatively narrow, showing diverging growth patterns around the time of puberty. See Figure 5.4.

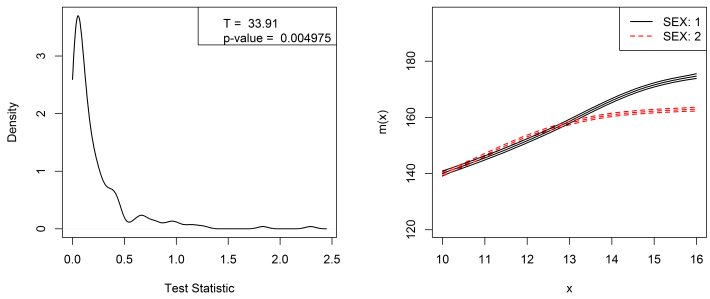

Figure 5.4: Empirical distribution of the test statistic under the null hypothesis and height over age by gender and pointwise 95% CI by gender

The height-for-age scatter plot is presented in Figure 5.4. The corresponding regression curves and the 95% pointwise confidence intervals from the semiparametric mixed model analysis are presented in Figure 5.4.

To compare surfaces we conduct hypothesis testing to determine whether the average simultaneous non-linear effects of height and age on weight differ by sex among the subset of white children only.

```
R> child.repw <- child.rep[(child.rep$RACE==1),]
R> child.reptest2 <- gamm4.grptest(WEIGHT~t2(age,HEIGHT),
      random=~(1|SID),test=~SEX,data=child.repw)
R> child.reptest2
R> plot(child.reptest2,type="contour")
R> plot(child.reptest2,type="persp",theta=-35,phi=40)
```

which produces the following output:

```
        Test the equality of surfaces based on L2 distance


Comparing 2 semiparametric regression surfaces
Penalized semiparametric regression mixed model is used for surface fitting.
Wide-bootstrap algorithm is applied to obtain the null distribution.
```

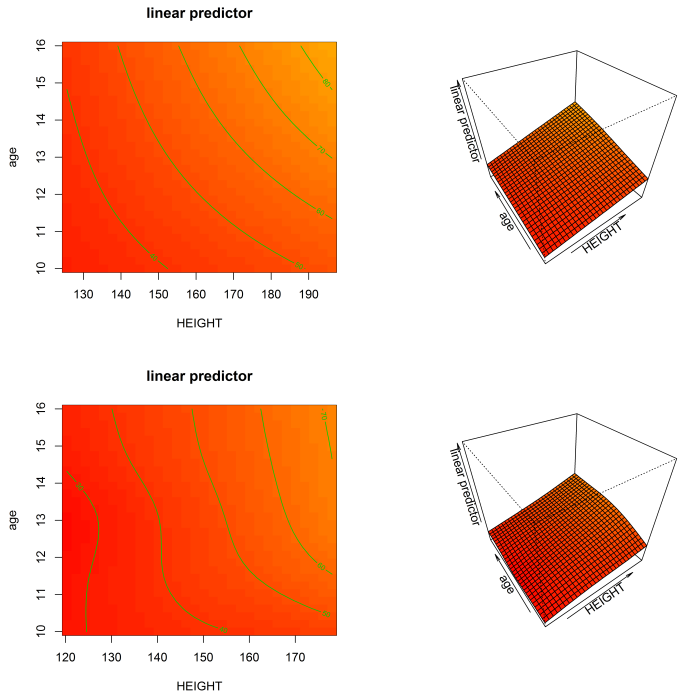

Figure 5.5: Estimated contour and 3D plots of weight on height and age by gender with correlated data

```
Null hypothesis: there is no difference between the 2 surfaces.

T =  10.88      p-value =  0.0995
```

As shown above, there is insufficient statistical evidence (p = 0.0995) to conclude that there was a difference in in the combined effect of weight and age by gender within white only children at significance level 0.05. Plots are shown in Figure 5.5. Setting argument `type=plotly.persp` generates an interactive 3-D plot using the R package `plotly`.

5.2 Interface by R Shiny

To enhance the usability of the testing methods, we created an interactive R Shiny interface for the `gamm4.test` package. This interface allows analysts that do not use R to

access the testing procedure through a web link. The interface can be access at `https://heather.shinyapps.io/shinygamm4/` and a youtube tutorial is available at `https://youtu.be/SHqaZXSLaMw`.

To illustrate, we use the `outchild` from "gamm4.test" as an example.

We first put the observed data in the prespecified format.

```
colnumes(childcur) <- c("grp", "y", "age")
```

Then click “Enter Data” to upload the dataset. To compare curves and plot the estimated regression functions, click "Test summary and plots".

For surface comparison, we first put the raw data in the specified format.

```
R> colnumes(childsurf) <- c("grp", "y", "x1", "x2")
```

We then compare surfaces by clicking ”Enter Data" and ”Test summary and plots".

BIBLIOGRAPHY

Beran, R. J. and G. R. Ducharme (1991). Asymptotic theory for bootstrap methods in statistics.

Bowman, A. (2006). Comparing nonparametric surfaces. *Statistical Modelling 6*, 279–299.

DasGupta, A. (2008). Asymptotic theory of statistics and probability.

de Boor, C. (2001). A practical guide to splines (revised edition).

Dette, H. and N. Neumeyer (2001). Nonparametric analysis of covariance. *The Annals of Statistics 29*(5), 1361–1400.

Draper, N. and H. Smith (1998). Applied regression analysis (third edition).

Eilers, P. and B. D. Marx (1996). Flexible smoothing with b–splines and penalties. *Statistical Science 11*(2), 89–121.

Eubank, R. (1999). Nonparametric regression and spline smoothing.

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *Journal of American Statistical Association 91*(434), 674–688.

Fan, J. and I. Gijbels (1996). Local polynomial modelling and its applications.

Fan, J. and S. Lin (1998). Test of significance when data are curves. *Journal of American Statistical Association 93*(443), 1007–1021.

Francisco-Fernandez, M. and J. Opsomer (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics 33*(2), 279–295.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 1218–1228.

Green, P. and B. Silverman (1994). Nonparametric regression and generalized linear models.

Gu, C. (2013). Smoothing spline anova models.

Hardle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 1926–1947.

Hardle, W., M. Muller, S. Sperlich, and A. Werwatz (2004). Nonparametric and semiparametric models.

Hart, J. D. (1997). Nonparametric smoothing and lack–of–fit tests.

Hastie, T. and R. Tibshirani (1990). Generalized additive models.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 255–285.

McMurry, T. L. and D. N. Politis (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis 31*(6), 471–482.

Nan, B., X. Lin, L. D. Lisabeth, and S. D. Harlow (1964). On estimating regression. *Theory of Probability and its Applications 9*(1), 141–142.

Neumeyer, N. and H. Dette (2003). Nonparametric comparison of regression curves: an empirical process approach. *The Annals of Statistics 31*(3), 880–920.

Ogden, R. (1996). Essential wavelets for statistical applications and data analysis.

Ruppert, D., M. Wand, and R. Carroll (2003). Semiparametric regression.

Ruppert, D., M. Wand, and R. Carroll (2009). Semiparametric regression during 2003-2007. *Electron J Stat 17*(1), 1193–1256.

Tu, W., G. Eckert, L. DiMeglio, Z. Yu, J. Jung, and J. Pratt (2011). Intensified effect of adiposity blood pressure in overweight and obese children. *Hypertension 58*, 818–824.

Tu, W., G. Eckert, T. Hannon, H. Liu, L. Pratt, M. Wagner, L. DiMeglio, J. Jung, and J. Pratt (2014). Racial differences in sensitivity of blood pressure to aldosterone. *Hypertension 58*, 1212–1218.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Wahba, G. (1990). Spline models for observational data.

Wand, M. and M. Jones (1995). Kernel smoothing.

Wang, X. and D. Ye (2010). On non–parametric comparison of images and regression surfaces. *J Stat Plan Inference 140*(10), 2875–2884.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A 26*(4), 359–372.

Wood, S. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B 65*(1), 95–114.

Wood, S. (2006). Low-rank scale-invariant tensor product smoothsfor generalized additive mixed models. *Biometrics 62*(4), 1025–1036.

Wood, S. (2018). mgcv 1.8-23. r package. `https://cran.r-project.org/web/packages/mgcv/index.html` (visited: 2018-05-09).

Young, S. G. and A. W. Bowman (1995). Non-parametric analysis of covariance. *Biometrics 51*, 920–931.

Zhang, D. and X. Lin (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics 4*(1), 57–74.

Zhang, D., X. Lin, J. Raz, and M. Sowers (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association 93*(1), 710–719.

Zhang, D., X. Lin, and M. F. Sowers (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics 56*(1), 31–39.

CURRICULUM VITAE

Shi Zhao

EDUCATION

- Ph.D. in Biostatistics - Indiana University (2018)
- M.S. in Applied Statistics, Human Nutrition - Ohio State University (2011)

WORKING EXPERIENCE

- Summer Intern, Vertex Pharmaceuticals Company (2016)
- Biostatistician, Medical College of Wisconsin (2011 - 2014)
- Summer Intern, Monsanto Company (2011)

SELECT PUBLICATIONS

- Shi Zhao, Giorgos Bakoyannis, Spencer Lourens, Wanzhu Tu. Statistical comparison of nonlinear curves and surfaces. Biometrics. (in progress)

PRESENTATIONS

- "Statistical comparison of nonlinear curves and surfaces". Department of Epidemiology and Biostatistics, Indiana University at Bloomington, March, 2018.
- "Statistical comparison of nonlinear curves and surfaces". Eastern North American Region (ENAR) Meeting, Atlanta, Georgia, March, 2018.
- "Visualizing and comparing nonlinear curves and surfaces". Midwest Biopharmaceutical Statistics Workshop, Indianapolis, Indiana, May, 2018.